

# Efficient Global Spatial-Angular Sparse Coding for Diffusion MRI with Separable Dictionaries

Evan Schwab, René Vidal, and Nicolas Charon

Center for Imaging Science, Johns Hopkins University, Baltimore, MD USA

---

## Abstract

Diffusion MRI (dMRI) can reconstruct neuronal fibers in the brain, *in vivo*, by measuring water diffusion along angular gradient directions in q-space. High angular resolution diffusion imaging (HARDI) can produce better estimates of fiber orientation than the popularly used diffusion tensor imaging, but the high number of samples needed to estimate diffusivity requires lengthy patient scan times. To accelerate dMRI, compressed sensing (CS) has been utilized by exploiting a sparse representation of the data, discovered through sparse coding. The sparser the representation, the fewer samples are needed to reconstruct a high resolution signal with limited information loss and so a focus of much dMRI research has been finding the sparsest possible dictionary representation. All prior methods, however, rely on an angular model of q-space signals in each voxel which fundamentally limits the global sparsity level since at least one dictionary atom is needed for each voxel. In contrast, we formulate a global spatial-angular representation of dMRI that will allow us to sparsely model an entire dMRI brain signal below the limit of one atom per voxel using joint spatial-angular sparse coding. But a main challenge is optimizing over large-scale dMRI data. In this work, we present extensions to a number of sparse coding algorithms that are better suited for large-scale problems by exploiting the separable Kronecker structure of our global spatial-angular dictionary. We compare the complexity and speed of our methods with prior Kronecker sparse coding algorithms and show promising sparsity results on phantom and real HARDI brain data for various dictionary choices. With great efficiency our method achieves significantly sparser HARDI representations than the state-of-the-art which has the potential achieve new levels of HARDI acceleration within a unified (k,q)-CS framework.

**Keywords:** diffusion MRI, HARDI, sparse coding, spatial-angular dictionaries, Kronecker, LASSO, OMP.

---

## 1. Introduction

Diffusion magnetic resonance imaging (dMRI) is a medical imaging modality used to analyze neuroanatomical biomarkers for brain diseases such as Alzheimer's. dMRI are 6D signals consisting of a 3D spatial MRI domain (k-space) and a 3D diffusion domain (q-space). In each voxel in a brain dMRI, 3D models of diffusion are reconstructed from the q-space measurements which are used to estimate orientations and integrity of neuronal fiber tracts, *in vivo*. Different dMRI protocols measure q-space in different ways. For example, diffusion spectrum imaging (DSI) [1] measures q-space densely on a 3D grid. Alternatively, diffusion tensor imaging (DTI) [2] simplifies acquisition by restricting q-space measurements to the unit q-sphere. High angular resolution diffusion imaging (HARDI) [3] also restricts measurements to the unit q-sphere, but increases the angular resolution from that of DTI. Multi-Shell HARDI (MS-HARDI) [4] expands its radial range to include multiple q-spheres, or shells. Since DTI collects the fewest number of measurements, it has become the most widely used clinical dMRI protocol. However, its simple tensor model is unable to capture the complex diffusion profiles in each voxel. On the other hand, protocols like HARDI, MS-HARDI, and especially DSI, collect a higher number of q-space measurements to estimate more accurate diffusion profiles, but suffer from lengthy scan times which make them not currently suitable for clinical studies.

An ongoing research goal has been to find ways to reduce acquisition times of HARDI, MS-HARDI, or DSI, while maintaining accurate estimations of diffusion. One avenue is from a hardware perspective: maintain dense signal measurement configurations while devising faster physical acquisition techniques like simultaneous multi-slice acquisition [5] and simultaneous image refocusing [6]. The other is from a signal processing perspective: maintain accurate signal reconstructions while devising methods to exploit redundancies in the

data to reduce the number of required measurements and accelerate acquisition. Eventually, the two areas may be optimally integrated [7], but the focus of this work is on advancing the latter through the use of the well-studied Compressed Sensing (CS) paradigm [8].

CS, a class of mathematical results and algorithms showing that by exploiting sparse representations of data, signals can be acquired at sub-Nyquist rates and reconstructed with extremely limited loss of accuracy, has been classically used to subsample k-space to accelerate structural MRI [9]. CS has already been applied extensively to accelerate dMRI by subsampling q-space (q-CS) [10] and more recently extended to subsample both k- and q-space ((k,q)-CS) [11]. However, dMRI signals are traditionally reconstructed per-voxel in order to estimate profiles of diffusivity at each voxel. Accordingly, prior work in CS for dMRI have limited their sparse coding to the angular domain by reconstructing sparse angular representations of q-space signals in each voxel. Therefore, at least one dictionary atom would be required to represent signals in each voxel and so the global sparsity over the whole dMRI dataset could be no less than the number of voxels.

The reduction of signal measurements is proportional to the level of sparsity of the data representation, and therefore state-of-the-art dMRI CS methods may have a fundamental limit to the accelerations that are possible by using angular data representations per-voxel. To overcome this limitation, we consider a framework that allows global sparsity levels below one atom per voxel, opening up the possibility to achieve much higher acceleration rates of dMRI than prior models. In this work, we propose a spatial-angular sparse coding of dMRI using a global spatial-angular representation which exploits redundancies in the spatial and angular domains, *jointly*. Incorporating this spatial-angular sparse coding framework into a unified (k,q)-CS framework will be the work of the sequel. A major challenge, however, is the computational complexity of performing a massive global optimization over large-scale dMRI data. Therefore our main contribution in this paper is to provide and compare several computationally efficient algorithms to solve the large-scale global spatial-angular coding problem for dMRI. We derive novel adaptations of known sparse coding algorithms by exploiting the separability of our spatial-angular dictionary. We show on phantom and real HARDI brain data that it is possible to achieve accurate global HARDI reconstructions with a sparse representation of less than one dictionary atom per voxel, exceeding the theoretical limit of the state-of-the-art.

The remainder of this paper is organized as follows: In Section 2, we provide a review of state-of-the-art CS methods in dMRI and illustrate the limitations of their performance on a phantom HARDI dataset. In Section 3, we present our global spatial-angular HARDI representation and formalize the spatial-angular sparse coding problem. Then, in Section 4, we develop and compare a set of novel algorithms to efficiently solve our large-scale global optimization. Finally, in Section 5 we provide experimental results showing the performance of our method over the state-of-the-art and conclude with a discussion in Section 6.

## 2. State-of-the-Art

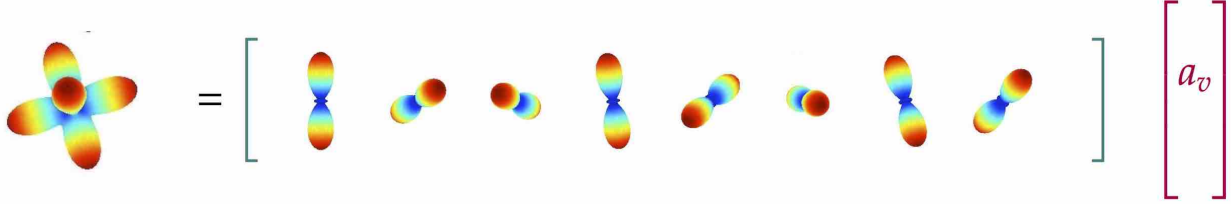
### 2.1. Angular (Voxel-Based) Reconstruction

A dMRI is a 6D signal  $\mathcal{S}(v, \vec{q}) \in \Omega \times \mathbb{R}^3$ , consisting of a set of  $G$  diffusion weighted images (DWIs) in 3D spatial domain  $\Omega \subset \mathbb{R}^3$  with  $V$  voxels, each acquired at a point in 3D q-space  $\{\vec{q}_g\}_{g=1}^G \in \mathbb{R}^3$ . An equivalent interpretation is that for each voxel  $v$  in a 3D brain volume  $\Omega$ , a vector of  $G$  diffusion measurements  $\{s_v(\vec{q}_g)\}_{g=1}^G$  are acquired at points  $\vec{q}_g$  in 3D q-space. The latter interpretation is most common for modeling because a major goal of dMRI reconstruction is to estimate 3D probability distribution functions (PDFs) of fiber tract orientation at each voxel. Accordingly, the signal vector  $s_v$  is represented by a q-space basis,  $\{\Gamma_i(\vec{q})\}_{i=1}^{N_\Gamma}$ , with  $N_\Gamma$  atoms, such that

$$s_v(\vec{q}) = \sum_{i=1}^{N_\Gamma} \Gamma_i(\vec{q}) a_{i,v}. \quad (1)$$

The dMRI literature has produced a wide array of dMRI reconstruction algorithms for different acquisition protocols, an artillery of q-space bases and varying models for estimating fiber tract PDFs. The vast majority of research reconstructs q-space signals in each voxel with a q-space basis while adding a set of constraints  $\mathcal{C}(a_v)$  on the coefficients to enforce desirable properties such as non-negativity of PDFs [12, 13], smoothing [14] or spatial coherence [15], solving:

$$a_v^* = \arg \min_{a_v} \frac{1}{2} \|\Gamma a_v - s_v\|_2^2 \quad \text{s.t.} \quad \mathcal{C}(a_v). \quad (2)$$



**Figure 1:** Illustration of per-voxel HARDI representation using a sparsifying dictionary.

The constraint of particular interest in our paper is that of enforcing sparsity on the coefficients of the reconstruction, known as *Sparse Coding*, which has applications in CS as well as super-resolution [16] and de-noising [17].

### 2.2. Angular (Voxel-Based) Sparse Coding

Sparse coding is a reconstruction problem which seeks a sparse solution, i.e. a coefficient vector  $a_v$  with very few nonzero elements. Given a sparsifying q-space basis  $\Gamma$  for which the dMRI signal in each voxel is expected to have a sparse representation, the angular (voxel-wise) sparse coding problem can be formulated as:

$$a_v^* = \arg \min_{a_v} \frac{1}{2} \|\Gamma a_v - s_v\|_2^2 \quad \text{s.t.} \quad \|a_v\|_0 \leq K_v. \quad (3)$$

where  $\|a_v\|_0$  counts the number of nonzero elements of vector  $a_v$ , and  $K_v$  is the sparsity level at voxel  $v$ . This problem is known to be NP-hard, and therefore the two main methodologies to tackle (3) are a) approximate a solution using greedy algorithms such as Orthogonal Matching Pursuit (OMP) [18] or b) relax (3) with the  $l_1$  norm using Basis Pursuit or solve the LASSO problem:

$$a_v^* = \arg \min_{a_v} \frac{1}{2} \|\Gamma a_v - s_v\|_2^2 + \lambda \|a_v\|_1, \quad (4)$$

using algorithms such as Alternating Method of Multipliers (ADMM) [19] or Fast Iterative Thresholding Algorithm (FISTA) [20], where  $\lambda$  is the trade-off parameter between data fidelity and sparsity. Angular sparse coding and q-space CS (q-CS) have been widely researched for dMRI to reduce long acquisition times. Many groups have done extensive work choosing sparsifying q-space bases [21, 10, 22], developing dictionary learning methods [23, 24, 25, 26, 27, 28, 29], and testing q-space subsampling schemes for DSI [30, 31, 32, 27, 33], MS-HARDI [34, 35, 25, 36, 37] HARDI [38, 39, 40, 41, 42] and DTI [43] with promising results in sparsity and measurements reduction for clinical tractography use [44, 45]. However, a major limitation for this set of prior work is that the sparsest possible representation of an entire dMRI dataset can be no less than the number of voxels since  $\|a_v\|_0 \geq 1 \quad \forall v \in \Omega$ . Therefore, the amount of q-space measurement reduction using CS for these methods are fundamentally restricted. In practice, the global sparsity level will be much greater than the number of voxels to account for noise. For example the work of [46, 39] report an average sparsity level  $K_v = 6$  to 10 atoms, i.e. a global sparsity level of at least  $6V$ , for HARDI using a spherical ridgelet dictionary. In addition, these methods do not attempt to exploit spatial redundancies in neighboring voxels and measurement reductions in k-space (k-CS).

### 2.3. Angular Sparse Coding with Spatial Regularization

Some works have extended angular sparse coding to include spatial regularization of neighboring voxels to improve reconstruction accuracy at consistent sparsity levels. While some [17, 47, 14, 48] have applications of de-noising or tractography, for which minimizing sparsity levels may not be the goal, others [49, 50, 51, 52, 53] use spatial regularization to exploit additional redundancies for q-CS. To further reduce dMRI measurements, the recent work of [54, 7, 55, 56, 57] combine k- and q-CS by adding a spatial sparsity term and a subsampled spatial data fidelity term. The work of [58, 11, 59, 60, 61] exploit subsampling in both k- and q-space but fail to include a sparsifying dictionary for the spatial domain which will limit the extent of measurement reductions. For example, the work of [55] naturally perform sparse MRI plus sparse dMRI, combining SparseMRI [9] for the spatial domain and previous formulations for the angular domain. Table 1 summarizes the recent

		Sparse Coding			
CS		Spatial	Angular	Spatial + Angular	Joint Spatial-Angular
	k	[62] <sup>†</sup>	[57]		
	q		[33, 53, 10, 29, 37, 22, 42, 48]		
	k + q		[11, 59, 60, 61]	[58, 7, 55, 56]	
	(k,q)				Proposed*

**Table 1:** Summary of all dMRI state-of-the-art from 2015 to present organized by domains of sparse coding and CS subsampling. The literature has provided a natural extension from spatial sparse coding in MRI and k-CS to angular sparse coding in dMRI and q-CS. However, when moving to (k,q)-CS, the most natural extension is the proposed joint spatial-angular sparse coding. (\*Though our proposed framework is intended for (k,q)-CS, the work in this paper performs joint spatial-angular sparse coding and not yet (k,q)-CS. <sup>†</sup>MRI reference is not exhaustive.)

literature’s usages of sparse coding and CS for dMRI. Each of the aforementioned methods aims to solve a program of the form

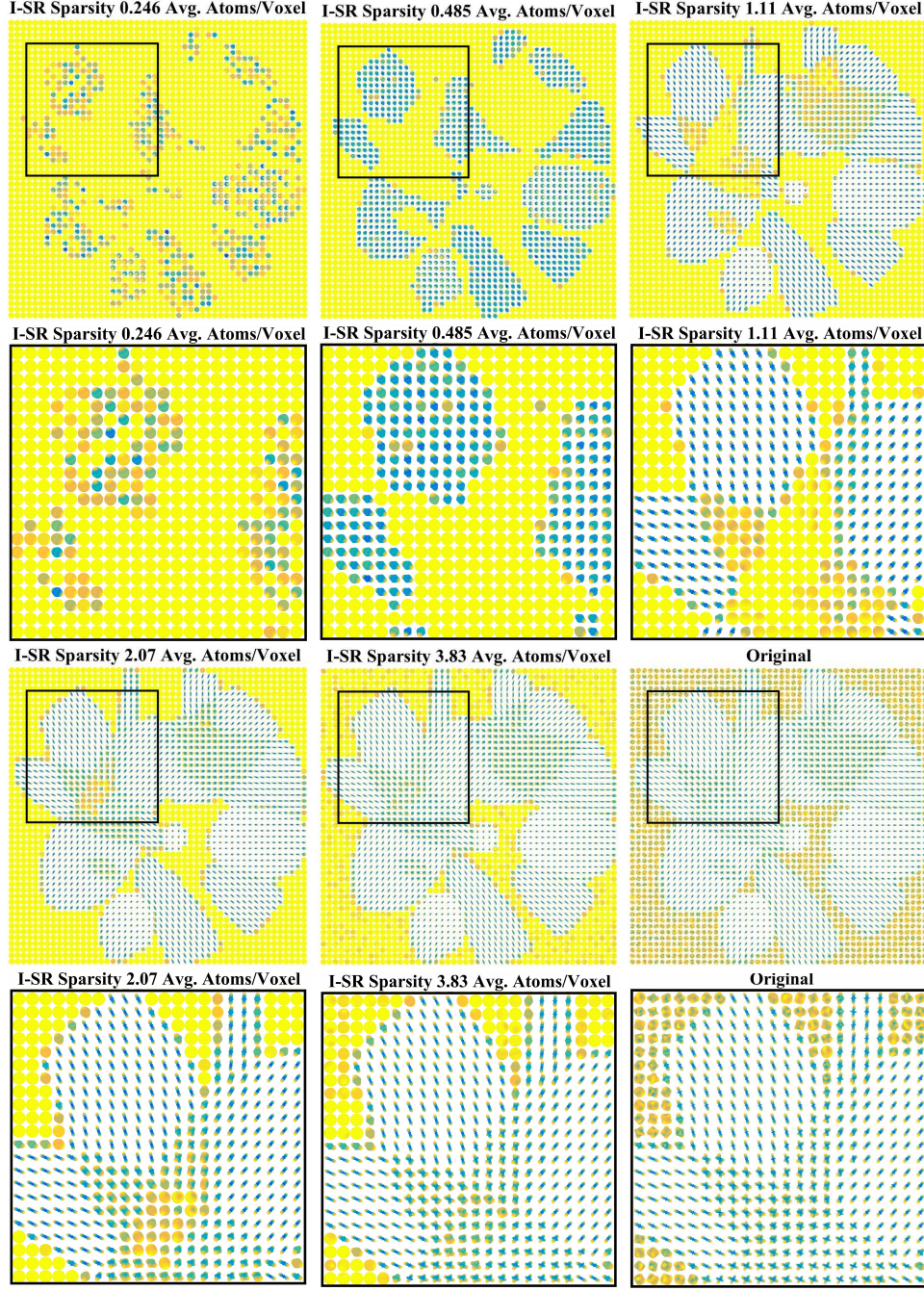
$$A^* = \arg \min_A \|\Gamma A - S\|_F^2 + \lambda \|A\|_1 + \mathcal{R}(A), \quad (5)$$

where  $S = [s_1 \dots s_V] \in \mathbb{R}^{G \times V}$  is the concatenation of signals  $s_v \in \mathbb{R}^G$  sampled at  $G$  gradient directions over  $V$  voxels,  $A = [a_1 \dots a_V] \in \mathbb{R}^{N_\Gamma \times V}$  is the concatenation of coefficients and  $\mathcal{R}(A)$  is a spatial regularizer, spatial sparsity term or subsampled spatial data fidelity term that depend on the angular representation  $A$ . In total, each of these formulations are based on an angular representation and reconstruction of dMRI data. In particular, additional spatial sparsity terms to (5) may exploit redundancies in both angular and spatial domains but because they are disjoint terms the minimal global sparsity level is still limited by the number of voxels. Furthermore, because the dMRI data is reconstructed with only a sparsifying angular dictionary, sub-sampling in k- and q-space does not exactly follow the theoretical setting of classical CS, for which theories relating sparsity and coherence provide guarantees of measurement reduction.

#### 2.4. State-of-the-Art Results

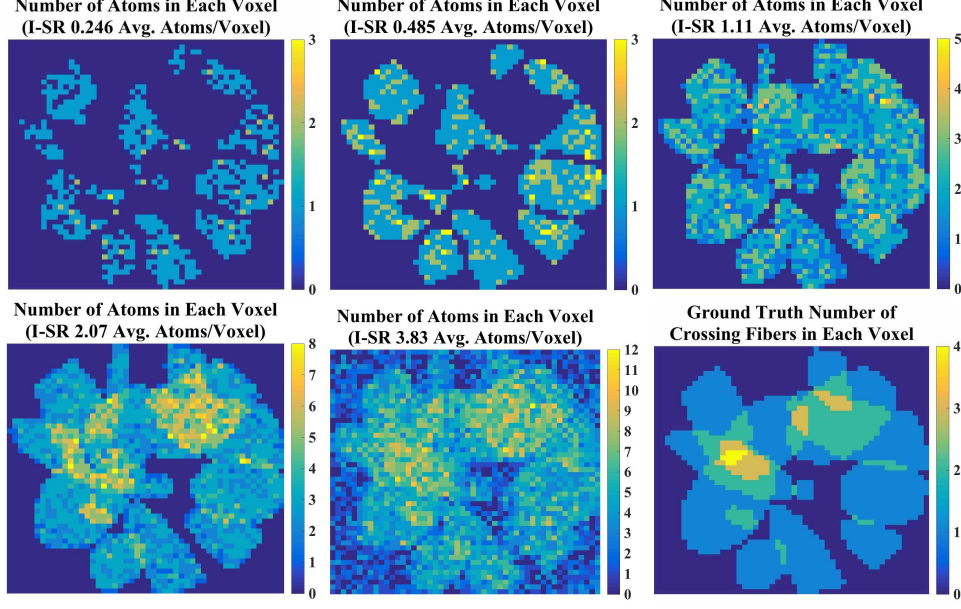
In Figure 2 we demonstrate the limitation of sparsity for state-of-the-art methods. In particular, we solve (5) with  $\mathcal{R} = 0$ , which is the equivalent of solving (4) for all voxels simultaneously. In contrast to (4), the global sparsity level  $K$  can actually fall below the number of voxels. However, this will result in some voxels (columns of  $A$ ) being represented with zero atoms, meaning that there exists no signal in that voxel, which is not physically representative of real dMRI data. For angular basis  $\Gamma$  we choose the well performing spherical ridgelet (SR) dictionary [46, 38, 40] on a  $V = 50 \times 50$  voxel 2D slice of HARDI phantom data with  $G = 64$  gradient directions. We solve (5) for 5 different values of  $\lambda$  which gives us global sparsity levels of 0.246, 0.485, 1.11, 2.07, and 3.84 average dictionary atoms per voxel ( $K/V$ ), averaged over the whole 2D image. We compare the state-of-the-art sparse reconstruction  $\hat{S}$  to the original signal  $S$  using an average residual,  $0.5\|\hat{S} - S\|_F^2/GV$ . We can see that for high global sparsity levels, the state-of-the-art method is unable to model crossing fibers. Figure 3 shows the number of atoms used to represent the HARDI signals in each voxel and the distribution of nonzero atoms in the entire image. The bottom right image shows the ground truth number of fibers crossing in each voxel. This demonstrates that voxels containing crossing fibers are forced to zero atoms for high sparsity levels and require 6-12 atoms for lower sparsity levels which is consistent with the reports of [46, 39] for SR.

In contrast to all prior work on sparse coding and CS in dMRI, which at its core relies on a per-voxel angular representation of dMRI data, we present in the following section a spatial-angular dictionary representation of dMRI from which we can apply joint spatial-angular sparse coding to exploit the redundancies in the spatial and angular domains *jointly*. With a global representation we allow for the possibility to achieve sparsity levels below the number of voxels without forcing voxels to have a zero atom representation. With a global spatial-angular dictionary representation, a unified joint (k,q)-CS will be implemented in future work.



**Figure 2:** Qualitative demonstration of state-of-the-art limitations using angular sparse coding optimization (5) with spherical ridgelets (SR) dictionary and  $\mathcal{R} = 0$  for 5 different global sparsity levels compared to original signal (bottom right) with ROI closeups underneath. For high global sparsity levels (top left, middle), voxels with complex signals are forced to use zero atoms. Regions with crossing fibers are unable to be accurately reconstructed even up to lower global sparsity levels of 2.07 avg. atoms/voxel, showing limitation of state-of-the-art sparsity levels. The label I-SR refers to Identity-SR, explained in the next section.





**Figure 3:** First 5 figures show the number of atoms in each voxel found by angular sparse coding with the spherical ridgelet dictionary (SR) for the 5 levels global sparsity presented above in Figure (2). The final figure shows the ground truth number of fibers crossing in each voxel to illustrate the complexity of each angular signal in relation to how many atoms are required to sparsely model them. Crossing fiber signals are either forced to zero for high global sparsity levels (see: top row) or require between 3-5 atoms for single fiber signals (see: avg. sparsity 1.11 and 2.07) and 6-12 for double and triple crossing fiber signals (see: avg. sparsity 3.83). The label I-SR refers to Identity-SR, explained in the next section.

### 3. Global Spatial-Angular dMRI Representation

In this work, we propose to model a dMRI signal  $\mathcal{S}(v, \vec{q}) \in \Omega \times \mathbb{R}^3$  globally with a joint spatial-angular representation, say  $\varphi(v, \vec{q}) \in \Omega \times \mathbb{R}^3$ , such that

$$\mathcal{S}(v, \vec{q}) = \sum_k c_k \varphi_k(v, \vec{q}) \quad (6)$$

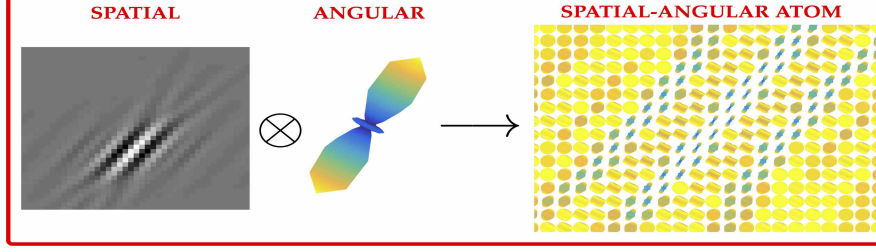
to explicitly reduce redundancies in both the spatial and angular domains jointly with a single set of globally sparse coefficients  $c = [c_k]$ . We let  $s \in \mathbb{R}^{VG}$  be the vectorization of  $\mathcal{S}(v, \vec{q})$  and  $\phi_k \in \mathbb{R}^{VG}$  be the vectorization  $\varphi_k(v, \vec{q})$  to build the spatial-angular matrix  $\Phi = [\phi_1 \dots \phi_{N_\Phi}] \in \mathbb{R}^{VG \times N_\Phi}$ , where  $N_\Phi$  are the number of atoms of the global dictionary. Then, to find a globally sparse  $c$ , we can approximate the  $l_0$  minimization problem:

$$c^* = \arg \min_c \frac{1}{2} \|\Phi c - s\|_2^2 \quad \text{s.t.} \quad \|c\|_0 \leq K, \quad (P0vec)$$

or solve the  $l_1$  minimization problem:

$$c^* = \arg \min_c \frac{1}{2} \|\Phi c - s\|_2^2 + \lambda \|c\|_1, \quad (P1vec)$$

where  $\lambda > 0$  is the sparsity trade-off parameter. However, typical dMRI contains on the order of  $V \approx 100^3$  voxels each with  $G \approx 100$  q-space measurements for a total of  $100^4 = 100$  million signal measurements ( $|s| \approx 10^8$ ). Since many sparse coding applications often utilize dictionaries that are over-redundant, this leads to a massive matrix  $\Phi$  with  $100^4$  rows and over  $100^4$  columns ( $|\Phi| \approx 10^{16}$ ). For some datasets, even committing  $\Phi$  to memory is prohibitive. Therefore solving this large-scale global dMRI sparse coding problem using traditional solvers like OMP to approximate (P0vec) or ADMM and FISTA to solve (P1vec), prove intractable.



**Figure 4:** Example construction of a single spatial-angular basis atom  $\Phi_k$  (right) by taking the Kronecker product of  $\Psi_j$  (left) and  $\Gamma_i$  (middle), i.e.  $\Psi_j \otimes \Gamma_i = \Phi_k$ . With this particular combination of spatial (Curvelet [63]) and angular (Spherical Wavelet [40]) atoms, we can see that it may be possible to represent an entire fiber tract with very few spatial-angular atoms.

To address this challenge, we introduce additional structure on the dictionary atoms by considering separable functions over  $\Omega$  and  $\mathbb{R}^3$ , namely a set of atoms of the form  $\varphi(v, \vec{q})_{i,j} = (\psi_j(v)\gamma_i(\vec{q}))$ , where  $\psi(v)$  is a spatial basis for  $\Omega$  and  $\gamma(\vec{q})$  is an angular basis for  $\mathbb{R}^3$ . The dMRI signal may then be decomposed as:

$$\mathcal{S}(v, \vec{q}) = \sum_k c_k \varphi_k(v, \vec{q}) = \sum_{i=1} \sum_{j=1} c_{i,j} \psi_j(v) \gamma_i(\vec{q}), \quad (7)$$

A motivating model for this separable structure is as follows: First, as is traditionally done, the signal at each voxel  $v \in \Omega$  is written as a linear combination of angular basis function  $\gamma$ :

$$S(v, \vec{q}) = \sum_{i=1} a_i(v) \gamma_i(\vec{q}). \quad (8)$$

Then, we notice that each spherical coefficient  $a_i(v)$  forms a 3D volume and so can be written as a linear combination of spatial basis function  $\psi$ :

$$a_i(v) = \sum_{j=1} c_{i,j} \psi_j(v). \quad (9)$$

Then combining (8) and (9) we arrive at our proposed separable spatial-angular representation (7). In discretized form, our global basis  $\varphi$  is the separable Kronecker product matrix  $\Phi := \Psi \otimes \Gamma$  with  $\Psi \in \mathbb{R}^{V \times N_\Psi}$  and  $\Gamma \in \mathbb{R}^{G \times N_\Gamma}$  and  $N_\Phi = N_\Psi N_\Gamma$  such that

$$s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_V \end{pmatrix} = \begin{pmatrix} \Psi_{1,1}\Gamma & \Psi_{1,2}\Gamma & \cdots & \Psi_{1,N_\Psi}\Gamma \\ \Psi_{2,1}\Gamma & \Psi_{2,2}\Gamma & \cdots & \Psi_{2,N_\Psi}\Gamma \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{V,1}\Gamma & \Psi_{V,2}\Gamma & \cdots & \Psi_{V,N_\Psi}\Gamma \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{N_\Psi N_\Gamma} \end{pmatrix} = \Phi c. \quad (10)$$

Then we explicitly exploit the separability of  $\Phi$  by turning to the equivalent matrix from of (10):

$$S = \Gamma C \Psi^\top. \quad (11)$$

This formulation allows us to avoid the expensive uses of  $\Phi$  and fully reduce computational complexity to the smaller separable basis domains of  $\Gamma$  and  $\Psi$ . The equivalent spatial-angular sparse coding problem statements we wish to solve are:

$$C^* = \arg \min_C \frac{1}{2} \|\Gamma C \Psi^\top - S\|_F^2 \quad \text{s.t.} \quad \|C\|_0 \leq K \quad (P0mat)$$

and

$$C^* = \arg \min_C \frac{1}{2} \|\Gamma C \Psi^\top - S\|_F^2 + \lambda \|C\|_1, \quad (P1mat)$$

where the  $l_0$  and  $l_1$  norms are the usual norms taken over all elements of the matrix  $C$ . Therefore, all operations will be computed in the considerably smaller dimensional spaces of  $\Gamma$  and  $\Psi$ . As an important

note, in the special case of  $\Psi = I_V$ , the  $V \times V$  identity matrix, we can see this leads to the state-of-the-art voxel-based formulation (5) with  $C \equiv A$ . In the Experiments Section 5 we will use this when comparing the performance of purely angular sparse coding with our proposed framework.

Decomposing signals into Kronecker (or more general multi-tensor) structures has been well researched to exploit separability for sparse coding [64], dictionary learning [65] and CS [66]. Our model fits comfortably into this setting, but heretofore, these methods have not considered the specific structure and size of dMRI data. Only [67] has applied multi-tensor sparse coding methods on dMRI data, but for the application of data compression for fiber tractography. In the next section, we will develop a set of algorithms to solve ( $P0mat$ ) and ( $P1mat$ ) efficiently for large-scale dMRI data.

## 4. Efficient Kronecker Sparse Coding Algorithms

In what follows we present three novel adaptations of existing sparse coding algorithms for solving large-scale sparse coding problems with a Kronecker dictionary structure. These are Kronecker extensions of OMP, ADMM, and FISTA. We compare these to an existing Kronecker sparse coding algorithms, a Kronecker OMP [68] as well as Kronecker Dual ADMM, developed in our prior work [69] and derived in a new formulation for comparison in this paper. We compare these algorithms in terms of complexity for various types of bases in Section 4.5 and show experimental time comparisons in Section 5.

### 4.1. Kronecker OMP

#### 4.1.1. Prior Work

To approximate a solution to the  $l_0$  problem ( $P0vec$ ), Orthogonal Matching Pursuit (OMP) [18] is a popular greedy algorithm that iteratively selects the atom that is most correlated with the signal, orthogonalizes it to the previously selected atoms by solving a least squares optimization, and selects the next atom that is most correlated with the resulting residual. For the case of a Kronecker structured basis, a Kronecker OMP (Kron-OMP) algorithm has been previously proposed [68, 64] that reduces computations of solving the least squares subproblem in each iteration by exploiting properties of the Kronecker product. This form of Kron-OMP, however, represents the signal, coefficients, and basis atoms in vector form providing a solution to ( $P0vec$ ). In Algorithm 1 we rewrite the Kron-OMP algorithm adapted to the structure of our problem. The main complexity gain of Kron-OMP over the vector OMP is the separating of  $\Gamma$  and  $\Psi$  when computing the maximally correlated atoms with the residual,  $|\Gamma^\top R \Psi|$  (See Alg. 1 Step 1) with complexity  $O(N_\Gamma G V + G N_\Psi N_\Psi)$  instead of computing  $|\Phi^\top r|$  with complexity  $O(N_\Gamma N_\Psi G V)$ . The other gain is in solving the least squares problem (See Alg 1 Step 3) by exploiting properties of the Kronecker product ( $A \odot B = [a_1 \otimes b_1, \dots, a_N \otimes b_N]$ ) to compute a rank-1 update. However, the only real improvement on complexity is in memory since  $\Phi$  can be built atom by atom from columns of  $\Gamma$  and  $\Psi$  instead of storing the entire matrix. The rank-1 update remains  $O(k^2)$  for both vector and Kron-OMP. In the next section we present an alternative Kron-OMP algorithm that reduces complexity further by exploiting the full separability of the dictionary.

---

#### Algorithm 1 Kron-OMP

---

Choose:  $K, \epsilon$ .  
Initialize:  $k = 1$ ,  $\mathcal{I}^0 = \emptyset$ ,  $\mathcal{J}^0 = \emptyset$ ,  $R_0 = S$ ,  $s = \text{vec}(S)$ .  
**while**  $k \leq K$  and error  $> \epsilon$  **do**  
  1:  $[i^k, j^k] = \arg \max_{[i,j]} |\Gamma^\top R_k \Psi|$ ;  
  2:  $\mathcal{I}^k = [\mathcal{I}^{k-1}, i^k]$ ;  $\mathcal{J}^k = [\mathcal{J}^{k-1}, j^k]$ ;  $\mathcal{A}^k = (\mathcal{I}^k, \mathcal{J}^k)$ ;  
  3:  $c_k = \arg \min_c \frac{1}{2} \|(\Gamma_{\mathcal{I}^k} \odot \Psi_{\mathcal{J}^k})c - s\|_2^2$ ;  
  4:  $R_k = \text{mat}(s - (\Gamma_{\mathcal{I}^k} \odot \Psi_{\mathcal{J}^k})c_k)$ ;  
  5:  $k \leftarrow k + 1$ ;  
**end while**  
Return:  $\mathcal{A}^K, c_K$

---



#### 4.1.2. Kronecker OMP with Projected Gradient Descent

In what follows, we develop a novel form of Kronecker OMP which solves the separable (*P0mat*) instead of (*P0vec*). This allows us to reduce computation by not building columns of  $\Phi$  and not repeating individual atoms of  $\Gamma$  or  $\Psi$ . Instead, indices of  $\Gamma$  and  $\Psi$  are updated only when they each have not been chosen before, exploiting the full separability of  $\Phi$ . Given the previous sets of respective of indices  $\mathcal{I}^{k-1}$  and  $\mathcal{J}^{k-1}$ , we update sets by following  $\mathcal{I}^k = [\mathcal{I}^{k-1} \ i^k]$  if  $i^k \notin \mathcal{I}^{k-1}$  and  $\mathcal{I}^k = \mathcal{I}^{k-1}$  otherwise. Likewise,  $\mathcal{J}^k = [\mathcal{J}^{k-1} \ j^k]$  if  $j^k \notin \mathcal{J}^{k-1}$  and  $\mathcal{J}^k = \mathcal{J}^{k-1}$  otherwise. With the selected indices, the size of  $C_k$  will be  $|\mathcal{I}^k| \times |\mathcal{J}^k|$  instead of  $k \times k$ . To find for  $C_k$ , it seems natural to solve:

$$C_k = \arg \min_C \frac{1}{2} \|\Gamma_{\mathcal{I}^k} C \Psi_{\mathcal{J}^k}^\top - S\|_F^2. \quad (12)$$

But the solution  $C_k$  will contain possible nonzero coefficients that do not coincide with the chosen selection of indices since additional indices in all combinations of pairs between  $\mathcal{I}^k$  and  $\mathcal{J}^k$  will be updated in each iteration. This is problematic for the correctness of the algorithm when choosing the next single most correlated coefficient. Therefore we must enforce that these coefficients are zero:

$$C_k = \arg \min_C \frac{1}{2} \|\Gamma_{\mathcal{I}^k} C \Psi_{\mathcal{J}^k}^\top - S\|_F^2 \text{ s.t. } C_{i,j} = 0 \ \forall (i,j) \in \mathcal{O}^k. \quad (13)$$

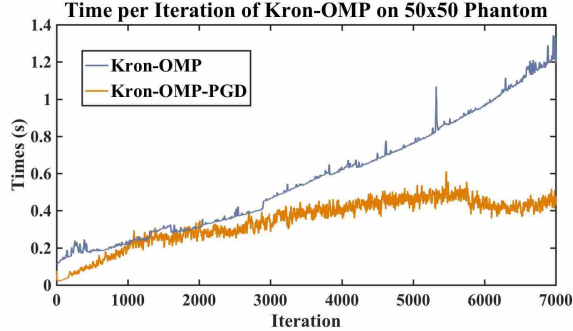
where  $\mathcal{O}^k := \overline{(\mathcal{I}^k, \mathcal{J}^k)}$ . To solve this problem, we can use projected gradient descent (PGD). The gradient of  $f(C) = \frac{1}{2} \|\Gamma_{\mathcal{I}^k} C \Psi_{\mathcal{J}^k}^\top - S\|_F^2$  at iteration  $k$  is

$$\nabla f(C) = \Gamma_{\mathcal{I}^k}^\top \Gamma_{\mathcal{I}^k} C \Psi_{\mathcal{J}^k}^\top \Psi_{\mathcal{J}^k} - \Gamma_{\mathcal{I}^k}^\top S \Psi_{\mathcal{J}^k}. \quad (14)$$

To save on computation we precompute  $\mathcal{G} = \Gamma^\top \Gamma$ ,  $\mathcal{P} = \Psi^\top \Psi$ , and  $\hat{S} = \Gamma^\top S \Psi$  and can access their entries at each iteration:  $\mathcal{G}_{\mathcal{I}^k, \mathcal{I}^k}$ ,  $\mathcal{P}_{\mathcal{J}^k, \mathcal{J}^k}$ ,  $\hat{S}_{\mathcal{I}^k, \mathcal{J}^k}$ . Then we iteratively project the update in the gradient direction to the space of feasible solutions:

$$C^{t+1} = P_{\mathcal{O}^k}(C^t - \epsilon \nabla f(C^t)) \quad (15)$$

where the projection  $P_{\mathcal{O}^k}$  sets all elements in  $\mathcal{O}^k$  to 0 and step-size  $\epsilon$  is estimated each iteration using a line search. Once the change in  $C^{t+1}$  and  $C^t$  is small enough, we arrive at our solution  $C_k = C^*$  and compute the residual  $R_k = S - \Gamma_{\mathcal{I}^k} C_k \Psi_{\mathcal{J}^k}^\top$ . Then, for iteration  $k+1$  we must find  $(i^{k+1}, j^{k+1}) = \arg \max_{[i,j]} |\Gamma^\top R_k \Psi|$ . To save significantly on computation we can instead use our precomputed  $\mathcal{G}$  and  $\mathcal{P}$  to instead find  $\arg \max_{[i,j]} |\hat{R}_k|$ , where  $\hat{R}_k = \hat{S} - \mathcal{G}_{\mathcal{I}^k} C_k \mathcal{P}_{\mathcal{J}^k}^\top$  where  $\mathcal{G}_{\mathcal{I}^k}$ ,  $\mathcal{P}_{\mathcal{J}^k}$  respectively access the  $\mathcal{I}^k$ ,  $\mathcal{J}^k$  columns and all rows. Maintaining matrix forms throughout allows us to combine computing the residual and the next atoms for a large reduction in computation at each iteration  $k$ . Our proposed Kronecker OMP with projected gradient descent (Kron-OMP-PGD) is outlined in Algorithm 2.



**Figure 5:** Comparison of time per iteration for Kron-OMP and the proposed Kron-OMP-PGD. The total time to choose  $K = 7000 = 2.8V$  atoms for this  $V = 50 \times 50$  slice of a phantom dataset, is 68 min for Kron-OMP and 40 min for Kron-OMP-PGD. We can see that as the number of atoms grows, the time per iteration of Kron-OMP continues to grow at a much higher rate than Kron-OMP-PGD.

We show a comparison of time per iteration for a small  $V = 50 \times 50, G = 64$  phantom dataset in Figure 5. The steeper time increase for Kron-OMP is due to the fact that at iteration  $k$  there is a complexity of  $O(k^2 + kGV)$  that comes from Steps 3 (rank-1 update) and 4 of Algorithm 1. On the other hand, Kron-OMP-PGD has complexity involving  $|\mathcal{I}^k|, |\mathcal{J}^k| \leq k$  which are in practice significantly less than  $k$ . Even though a PGD sub-routine must be performed at each iteration  $k$ , we found that by incorporating Nesterov acceleration with a line search, the time per iteration remains lower than Kron-OMP as the number of iterations  $k$  increases. However, for dMRI data, typical sparsity levels are  $K = O(V)$ . So for  $V \approx 100^3$  the number of iterations as well as the time per iteration of both Kron-OMP and Kron-OMP-PGD when  $k$  approaches  $K$  becomes astronomical. Even on a relatively small 3D phantom dataset of spatial size  $V = 50 \times 50$ , for example, one iteration takes on the order of a few seconds which results in over 34 hrs for these greedy algorithms to reach 1 atom/voxel atoms ( $K = V$ ). In this way, greedy algorithms such as OMP are not suitable for large-scale problems with large sparsity levels that require hundreds of thousands of iterations. The LASSO problem (*P1mat*) requires significantly less iterations and will be the topic of our next three algorithms.

---

**Algorithm 2** Kron-OMP-PGD

---

Choose:  $K, \epsilon_1, \epsilon_2$ .  
Precompute:  $\hat{S} = \Gamma^\top S \Psi$ ,  $\mathcal{G} = \Gamma^\top \Gamma$ ,  $\mathcal{P} = \Psi^\top \Psi$ .  
Initialize:  $k = 1$ ,  $\mathcal{I}^0 = \emptyset$ ,  $\mathcal{J}^0 = \emptyset$ ,  $\hat{R}_0 = \hat{S}$ .  
**while**  $k \leq K$  and error  $> \epsilon_1$  **do**  
  1:  $[i^k, j^k] = \arg \max_{[i,j]} |\hat{R}_k|$ ;  
  2:  $\mathcal{I}^k = \mathcal{I}^{k-1} \cup i^k$ ;  $\mathcal{J}^k = \mathcal{J}^{k-1} \cup j^k$ ;  $\mathcal{A}^k = (\mathcal{I}^k, \mathcal{J}^k)$ ;  $\mathcal{O}^k = \overline{\mathcal{A}^k}$ ;  
  3:  $Z_{\mathcal{J}^{k-1}, \mathcal{I}^{k-1}}^1 = C_{k-1}$ ;  $n_1 = 0$ ;  $t = 1$ ;  
  **while** error  $> \epsilon_2$  **do**  
    1:  $\delta = \text{linesearch}(Z^t)$ ;  
    2:  $C^{t+1} = P_{\mathcal{O}^k}(Z^t - \delta(\mathcal{G}_{\mathcal{I}^k, \mathcal{I}^k} Z^t \mathcal{P}_{\mathcal{J}^k, \mathcal{J}^k} - \hat{S}_{\mathcal{I}^k, \mathcal{J}^k}))$ ;  
    4:  $n_{t+1} = \frac{1}{2}(1 + \sqrt{1 + 4n_t^2})$ ;  
    5:  $Z^{t+1} = C^{t+1} + \frac{n_t - 1}{n_{t+1}}(C^{t+1} - C^t)$ ;  
    6:  $t \leftarrow t + 1$ ;  
  **end while**  
  4:  $C_k = C^*$ ;  
  5:  $\hat{R}_k = \hat{S} - \mathcal{G}_{\mathcal{I}^k} C_k \mathcal{P}_{\mathcal{J}^k}$ ;  
  6:  $k \leftarrow k + 1$ ;  
**end while**  
Return:  $\mathcal{A}^K, C_K$ .

---

#### 4.2. Kronecker ADMM

The Alternating Direction Method of Multipliers (ADMM) [19] is a popular method for solving the  $l_1$  minimization problem (*P1vec*), however, its application in the case of a large dictionary  $\Phi$  remains prohibitive, requiring computations involving  $\Phi^\top s$  of order  $O(GVN_\Gamma N_\Psi)$ . Instead, we apply ADMM to the separable  $l_1$  minimization problem (*P1mat*) to reduce computations by solving

$$\min_{C, Z} \frac{1}{2} \|\Gamma C \Psi^\top - S\|_F^2 + \lambda \|Z\|_1 \quad \text{s.t.} \quad C = Z. \quad (16)$$

The augmented Lagrangian writes:

$$\mathcal{L}_\mu(C, Z, \mathcal{T}) = \frac{1}{2} \|\Gamma C \Psi^\top - S\|_F^2 + \lambda \|Z\|_1 + \langle \mathcal{T}, C - Z \rangle + \frac{\mu}{2} \|C - Z\|_F^2, \quad (17)$$

and:

$$\frac{\partial \mathcal{L}_\mu}{\partial C} = \Gamma^\top (\Gamma C \Psi^\top - S) \Psi + \mathcal{T} + \mu(C - Z) = 0 \quad (18)$$

$$\implies \Gamma^\top \Gamma C \Psi^\top \Psi + \mu C = \mu Z - \mathcal{T} + \Gamma^\top S \Psi := Q \quad (19)$$

To solve for  $C$ , we begin by taking the SVDs of  $\Gamma$  and  $\Psi$ . With  $\Gamma = U_\Gamma \Sigma_\Gamma V_\Gamma^\top$  and  $\Psi = U_\Psi \Sigma_\Psi V_\Psi^\top$ ,  $\Gamma^\top \Gamma = V_\Gamma \Delta_\Gamma V_\Gamma^\top$  and  $\Psi^\top \Psi = V_\Psi \Delta_\Psi V_\Psi^\top$ , where  $U_\Gamma, U_\Psi$  are the matrices of eigenvectors and  $\Delta_\Gamma = \Sigma_\Gamma^\top \Sigma_\Gamma, \Delta_\Psi = \Sigma_\Psi^\top \Sigma_\Psi$  are the diagonal matrices of eigenvalues for  $\Gamma$  and  $\Psi$  respectively. Then:

$$V_\Gamma \Delta_\Gamma V_\Gamma^\top C V_\Psi \Delta_\Psi V_\Psi^\top + \mu C = Q \quad (20)$$

$$\implies \Delta_\Gamma \tilde{C} \Delta_\Psi + \mu \tilde{C} = \tilde{Q} \quad (21)$$

where we use the relationship  $\tilde{X} = V_\Gamma^\top X V_\Psi$ . Then, because  $\Delta_\Gamma$  and  $\Delta_\Psi$  are diagonal with elements  $\delta_{\Gamma_i}$  and  $\delta_{\Psi_j}$ , respectively, we can solve for  $\tilde{C}$  by:

$$\delta_{\Gamma_i} \tilde{C}_{i,j} \delta_{\Psi_j} + \mu \tilde{C}_{i,j} = \tilde{Q}_{i,j} \implies \tilde{C}_{i,j} = \frac{\tilde{Q}_{i,j}}{\delta_{\Gamma_i} \delta_{\Psi_j} + \mu} \quad (22)$$

To write this in matrix form we define  $\Sigma_\mu^{-1} \triangleq 1/(\delta_{\Gamma_i} \delta_{\Psi_j} + \mu)$  and have  $\tilde{C} = (\Sigma_\mu^{-1} \circ \tilde{Q})$  where  $\circ$  stands for element-wise matrix multiplication. Finally, we can recover  $C = V_\Gamma \tilde{C} V_\Psi^\top$ . In total, the update for  $C$  is:

$$C = V_\Gamma^\top (\Sigma_\mu^{-1} \circ (V_\Gamma Q V_\Psi^\top)) V_\Psi \quad (23)$$

where  $Q = \mu Z - \mathcal{T} + \Gamma^\top S \Psi$ . When minimizing  $\mathcal{L}_\mu$  with respect to  $Z$ , we end up with the usual proximal operator of the  $l_1$  norm that is given by the shrinkage operator,  $\text{shrink}_\kappa(X) = \max(0, X - \kappa) - \max(0, -X - \kappa)$ , applied element-wise to matrix  $X$ , such that  $Z_{k+1} = \text{shrink}_{\lambda/\mu}(C_{k+1} + \mathcal{T}_k)$ . For  $\frac{\partial \mathcal{L}_\mu}{\partial T}$ , we have the usual update in matrix form,  $\mathcal{T}_{k+1} = \mathcal{T}_k + C_{k+1} - Z_{k+1}$ . The formal updates for Kron-ADMM are presented in Algorithm 3. The update for  $C$  in (23) works well when  $\Gamma$  and  $\Psi$  are under-complete and the eigen-decompositions of

---

**Algorithm 3** Kron-ADMM (for undercomplete dictionaries)

---

Choose:  $\mu, \lambda, \epsilon$ .

Precompute:  $V_\Gamma, \Delta_\Gamma, V_\Psi, \Delta_\Psi, \Sigma_\mu^{-1}$ .

Initialize:  $k = 0, Z_0 = \mathbf{0}, \mathcal{T}_0 = \mathbf{0}$ .

**while** error  $> \epsilon$  **do**

- 1:  $Q = \Gamma^\top S \Psi + \mu(Z_k - \mathcal{T}_k)$ ;
- 2:  $C_{k+1} = V_\Gamma^\top (\Sigma_\mu^{-1} \circ (V_\Gamma Q V_\Psi^\top)) V_\Psi$ ;
- 3:  $Z_{k+1} = \text{shrink}_{\lambda/\mu}(C_{k+1} + \mathcal{T}_k)$ ;
- 4:  $\mathcal{T}_{k+1} = \mathcal{T}_k + C_{k+1} - Z_{k+1}$ ;
- 5:  $k \leftarrow k + 1$ ;

**end while**

Return:  $C^*$ .

---

$\Gamma^\top \Gamma$  and  $\Psi^\top \Psi$  are easily computable. However, dictionaries most commonly used for sparse coding and the application to CS are over-complete i.e.  $G < N_\Gamma$  and  $V < N_\Psi$  making these SVDs computationally expensive to calculate. In the case of an over-complete  $\Phi$ , for traditional vector ADMM, the matrix inversion lemma [19] is involved in order to compute SVDs of the smaller  $\Phi \Phi^\top$  instead of  $\Phi^\top \Phi$ . In the following proposition, we derive the equivalent result for the update of  $C$  in (23).

**Proposition 1.** *For over-complete dictionaries  $\Gamma$  and  $\Psi$ , update (23) is equivalent to the more compact*

$$C = Q/\mu - \Gamma^\top U_\Gamma (\Sigma_\mu^{-1} \circ (\Gamma U_\Gamma^\top Q U_\Psi \Psi^\top)) U_\Psi^\top \Psi / \mu. \quad (24)$$

*Proof.* For over-complete dictionaries  $\Gamma = U_\Gamma [\Sigma_\Gamma, \mathbf{0}] V_\Gamma^\top$  and  $\Psi = U_\Psi [\Sigma_\Psi, \mathbf{0}] V_\Psi^\top$ ,

$$\Gamma^\top \Gamma = V_\Gamma \begin{pmatrix} \Delta_\Gamma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} V_\Gamma^\top \text{ and } \Psi^\top \Psi = V_\Psi \begin{pmatrix} \Delta_\Psi & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} V_\Psi^\top.$$

For  $G < i \leq N_\Gamma, V < j \leq N_\Psi$ ,  $\delta_{\Gamma_i}, \delta_{\Psi_j} = 0$ , so  $\tilde{C}_{i,j} = \frac{\tilde{Q}_{i,j}}{\delta_{\Gamma_i}\delta_{\Psi_j} + \mu} = \frac{\tilde{Q}_{i,j}}{\mu}$ . For  $i \leq G$  and  $j \leq V$ , we can rewrite

$$\begin{aligned}\tilde{C}_{i,j} &= \frac{\tilde{Q}_{i,j}}{\delta_{\Gamma_i}\delta_{\Psi_j} + \mu} = \frac{\tilde{Q}_{i,j}}{\mu} - \frac{\delta_{\Gamma_i}\tilde{Q}_{i,j}\delta_{\Psi_j}}{\mu(\delta_{\Gamma_i}\delta_{\Psi_j} + \mu)} = \frac{\tilde{Q}_{i,j}}{\mu} - \frac{\sigma_{\Gamma_i}^2\tilde{Q}_{i,j}\sigma_{\Psi_j}^2}{\mu(\delta_{\Gamma_i}\delta_{\Psi_j} + \mu)} \\ &= \frac{\tilde{Q}_{i,j}}{\mu} - \sigma_{\Gamma_i}\frac{\sigma_{\Psi_j}\tilde{Q}_{i,j}\sigma_{\Psi_j}}{\mu(\delta_{\Gamma_i}\delta_{\Psi_j} + \mu)}\sigma_{\Psi_j} \\ \implies \tilde{C} &= \tilde{Q}/\mu - \Sigma_\Gamma^\top(\Sigma_\mu^{-1} \circ (\Sigma_\Gamma\tilde{Q}\Sigma_\Psi^\top))\Sigma_\Psi/\mu \\ C &= Q/\mu - V_\Gamma\Sigma_\Gamma^\top(\Sigma_\mu^{-1} \circ (\Sigma_\Gamma V_\Gamma^\top Q V_\Psi\Sigma_\Psi^\top))\Sigma_\Psi V_\Psi^\top/\mu \\ C &= Q/\mu - \Gamma^\top U_\Gamma(\Sigma_\mu^{-1} \circ (\Gamma U_\Gamma^\top Q U_\Psi\Psi^\top))U_\Psi^\top\Psi/\mu\end{aligned}$$

□

Letting  $\Gamma' = U_\Gamma^\top\Gamma$  and  $\Psi' = U_\Psi^\top\Psi$ , which can be precomputed, we have a final compact update

$$C = Q/\mu - \Gamma'^\top(\Sigma_\mu^{-1} \circ (\Gamma'Q\Psi'^\top))\Psi'/\mu \quad (25)$$

This allows us to compute the SVDs of  $\Gamma\Gamma^\top$  and  $\Psi\Psi^\top$  instead of the larger  $\Gamma^\top\Gamma$  and  $\Psi^\top\Psi$  and work with smaller matrices within each iteration. We present Kron-ADMM for over-complete dictionaries in Algorithm 4.

---

**Algorithm 4** Kron-ADMM (for overcomplete dictionaries)

---

Choose:  $\mu, \lambda, \epsilon$ .

Precompute:  $U_\Gamma, \Delta_\Gamma, U_\Psi, \Delta_\Psi, \Gamma', \Psi', \Sigma_\mu^{-1}$ .

Initialize:  $k = 0, Z_0 = \mathbf{0}, \mathcal{T}_0 = \mathbf{0}$ .

**while** error  $> \epsilon$  **do**

- 1:  $Q = \Gamma^\top S\Psi + \mu(Z_k - \mathcal{T}_k)$ ;
- 2:  $C_{k+1} = Q/\mu - \Gamma'^\top(\Sigma_\mu^{-1} \circ (\Gamma'Q\Psi'^\top))\Psi'/\mu$ ;
- 3:  $Z_{k+1} = \text{shrink}_{\lambda/\mu}(C_{k+1} + \mathcal{T}_k)$ ;
- 4:  $\mathcal{T}_{k+1} = \mathcal{T}_k + C_{k+1} - Z_{k+1}$ ;
- 5:  $k \leftarrow k + 1$ ;

**end while**

Return:  $C^*$ .

---

#### 4.3. Kronecker Dual ADMM

As an alternative to ADMM, Dual ADMM, which applies ADMM to the dual of  $l1$  problem (*P1vec*), has been shown to be more efficient than ADMM for over-complete dictionaries [70] by allowing one to compute SVDs of the more affordable  $\Phi\Phi^\top$  instead of  $\Phi^\top\Phi$ . In our previous work [69] we proposed a Kronecker Dual ADMM (Kron-DADMM) that efficiently solves the spatial-angular sparse coding problem. Below, we give an alternative derivation of this algorithm directly based on the matrix formulation of (*P1mat*). The dual of (*P1mat*) is:

$$\max_A -\frac{1}{2}\|A\|_F^2 + A^\top S \quad \text{s.t.} \quad \|\Gamma^\top A\Psi\|_\infty \leq \lambda. \quad (26)$$

To apply ADMM to this optimization problem, we replace  $\Gamma^\top A\Psi$  with auxiliary variable  $\mathcal{V}$  and add the additional constraint  $\Gamma^\top A\Psi - \mathcal{V} = 0$  to get:

$$\max_{A, \mathcal{V}} -\frac{1}{2}\|A\|_F^2 + A^\top S \quad \text{s.t.} \quad \|\mathcal{V}\|_\infty \leq \lambda \quad \text{and} \quad \mathcal{V} = \Gamma^\top A\Psi. \quad (27)$$

Then the augmented Lagrangian is

$$\mathcal{L}_\eta(A, \mathcal{V}, C) = -\frac{1}{2}\|A\|_F^2 + A^\top S + \langle C, \mathcal{V} - \Gamma^\top A\Psi \rangle + \frac{\eta}{2}\|\mathcal{V} - \Gamma^\top A\Psi\|_F^2 + \delta_\lambda(\mathcal{V}) \quad (28)$$

where

$$\delta_\lambda(\mathcal{V}) = \begin{cases} 0 & \text{if } \|\mathcal{V}\|_\infty \leq \lambda \\ \infty & \text{if } \|\mathcal{V}\|_\infty > \lambda \end{cases} \quad (29)$$

and  $C$  is the Lagrange multiplier corresponding the primal variable in  $(P1mat)$ . Then

$$\frac{\partial \mathcal{L}_\eta(A, \mathcal{V}, C)}{\partial A} = -A + S - \Gamma C \Psi^\top + \eta \Gamma (\mathcal{V} - \Gamma^\top A \Psi) \Psi^\top = 0 \quad (30)$$

$$\implies A + \eta \Gamma \Gamma^\top A \Psi \Psi^\top = S - \Gamma (C - \eta \mathcal{V}) \Psi^\top := P. \quad (31)$$

Now with eigen-decompositions  $\Gamma \Gamma^\top = U_\Gamma \Delta_\Gamma U_\Gamma^\top$  and  $\Psi \Psi^\top = U_\Psi \Delta_\Psi U_\Psi^\top$  and letting  $\tilde{X} = U_\Gamma^\top X U_\Psi$ ,

$$A + \eta U_\Gamma \Delta_\Gamma U_\Gamma^\top A U_\Psi \Delta_\Psi U_\Psi^\top = P \quad (32)$$

$$\implies \tilde{A} + \eta \Delta_\Gamma \tilde{A} \Delta_\Psi = \tilde{P}. \quad (33)$$

Then,  $\tilde{A}$  can be found element-wise by:

$$\tilde{A}_{i,j} + \eta \delta_{\Gamma_i} \tilde{A}_{i,j} \delta_{\Psi_j} = \tilde{P}_{i,j} \implies \tilde{A}_{i,j} = \frac{\tilde{P}_{i,j}}{1 + \eta \delta_{\Gamma_i} \delta_{\Psi_j}}. \quad (34)$$

Defining  $\Sigma_{\eta,j}^{-1} \triangleq 1/(1 + \eta \delta_{\Gamma_i} \delta_{\Psi_j})$ , the update is  $\tilde{A} = \Sigma_\eta^{-1} \circ \tilde{P}$ . As shown in [70] we can keep the update in terms of  $\tilde{A}$  instead of  $A$  since the variable we are interested in is  $C$ . We can then precompute  $S' = \Gamma'^\top S \Psi'$ ,  $\Gamma' = U_\Gamma^\top \Gamma$  and  $\Psi' = U_\Psi^\top \Psi$ . The updates of  $\mathcal{V}$  and  $C$  are as in [69] and presented in Algorithm 5, where  $P_\lambda^\infty(X)$  sets all entries of matrix  $X$  that are greater than  $\lambda$  to  $\lambda$ . The step-size  $\eta$  is kept constant to avoid the repeated computation of  $\Sigma_\eta^{-1}$

---

**Algorithm 5** Kron-DADMM

---

Choose:  $\eta, \lambda, \epsilon$ .

Precompute:  $S', \Gamma', \Psi', \Sigma_\eta^{-1}$ .

Initialize:  $k = 0, C_0 = 0, \mathcal{V}_0 = 0$ .

**while** Duality Gap  $> \epsilon$  **do**

1:  $\tilde{A}_{k+1} = \Sigma_\eta^{-1} \circ (S' - \Gamma' (C_k - \eta_k \mathcal{V}_k) \Psi'^\top)$ ;

2:  $\mathcal{V}_{k+1} = P_\lambda^\infty(\frac{1}{\eta} C_k + \Gamma'^\top \tilde{A}_{k+1} \Psi')$ ;

3:  $C_{k+1} = \text{shrink}_{\lambda\eta}(C_k + \eta \Gamma'^\top \tilde{A}_{k+1} \Psi')$ ;

4:  $k \leftarrow k + 1$ ;

**end while**

Return:  $C^*$ .

---

#### 4.4. Kronecker FISTA

The Fast Iterative Thresholding Algorithm (FISTA) [20] is another well-known method for solving LASSO. However, just as before, applying FISTA to  $(P1vec)$  for large-scale dMRI data is largely intractable. So here we apply FISTA to  $(P1mat)$  in order to exploit the separability of our spatial-angular basis. FISTA is a proximal gradient-descent

$$C_{k+1} = \text{shrink}_{\lambda/L}(C_k - \nabla f(C_k)/L), \quad (35)$$

where the proximal operator is the soft-thresholding shrinkage operator associated with the  $l1$  norm and  $1/L$  is a chosen step size. The gradient is simply computed as:

$$\nabla f(C) = \Gamma^\top (\Gamma C \Psi^\top) \Psi - \Gamma^\top S \Psi. \quad (36)$$

To help speed convergence, we invoke a line search subroutine to update  $L$  at each iteration in addition to the usual Nesterov acceleration. By [20], FISTA will converge for any  $L$  greater than the Lipschitz constant of  $\nabla f$  which can be estimated by bounding

$$\|\nabla f(C) - \nabla f(\bar{C})\|_F = \|\Gamma^\top \Gamma (C - \bar{C}) \Psi^\top \Psi\|_F \leq \lambda_{\max}^\Gamma \lambda_{\max}^\Psi \|C - \bar{C}\|_F \quad (37)$$



where  $\lambda_{\max}^{\Gamma}$  and  $\lambda_{\max}^{\Psi}$  are the maximum eigenvalues of  $\Gamma^{\top}\Gamma$  and  $\Psi^{\top}\Psi$  respectively. Therefore we initialize  $L = \lambda_{\max}^{\Gamma}\lambda_{\max}^{\Psi}$ . The Kronecker FISTA (Kron-FISTA) is presented in Algorithm 6. This natural Kronecker extension to FISTA has also been recently presented in [71], but has not been adapted and tested on data of our scale.

---

**Algorithm 6** Kron-FISTA

---

Choose:  $\epsilon$ .  
Precompute:  $\hat{S} = \Gamma^{\top}S\Psi$   
Initialize:  $Z_1 = C_0 = \mathbf{0}$ ,  $n_1 = 1$ ,  $L = \lambda_{\max}^{\Gamma}\lambda_{\max}^{\Psi}$ .  
**while** error  $> \epsilon$  **do**  
  1:  $L = \text{linesearch}(Z_k)$ ;  
  2:  $\nabla f(Z_k) = \Gamma^{\top}(\Gamma Z_k \Psi^{\top})\Psi - \hat{S}$ ;  
  3:  $C_k = \text{shrink}_{\lambda/L}(Z_k - \nabla f(Z_k)/L)$ ;  
  4:  $n_{k+1} = \frac{1}{2}(1 + \sqrt{1 + 4n_k^2})$ ;  
  5:  $Z_{k+1} = C_{k+1} + \frac{n_k - 1}{n_{k+1}}(C_{k+1} - C_k)$ ;  
  6:  $k \leftarrow k + 1$ ;  
**end while**  
Return:  $C^*$ .

---

#### 4.5. Complexity Analysis

To evaluate the efficiency of each algorithm and the gains of Kronecker separability compared to the original algorithms we summarize the complexity of each algorithm for general  $\Psi$  and  $\Gamma$  in Table 2. It appears that Kron-FISTA has the least complexity in the general setting. We compare the algorithms empirically in Section 5. Next we address the fact that the dimensions of  $\Gamma \in \mathbb{R}^{G \times N_{\Gamma}}$  and  $\Psi \in \mathbb{R}^{V \times N_{\Psi}}$  will be orders of magnitude different with  $G \approx 100$  and  $V \approx 100^3$ . We consider a few specific assumptions on the structure of spatial representation  $\Psi$  which can decrease the complexity and simplify computations of some of the proposed algorithms.

**$\Psi$  Tight Frame.** In the case that  $\Psi$  is a tight frame, *i.e.*  $\Psi\Psi^{\top} = I$ , which is commonly an assumption in CS theorems, our method can still be simplified. In Kron-ADMM (overcomplete) and Kron-DADMM, we may avoid the SVD of  $\Psi\Psi^{\top}$  and respective updates (21) and (33) can be simplified.

**$\Psi$  Fast Transform.** In the case that  $\Psi$  corresponds to a well-studied transform such as wavelets, curvelets, etc., fast transform implementations can be utilized to reduce complexity further. For the case of FISTA, for example, matrix multiplications of  $\Gamma^{\top}(\Gamma Z_k \Psi^{\top})\Psi$  involve fast transform reconstructions ( $\Psi^{\top}$ ) of each DWI ( $\Gamma Z_k$ ) and then deconstructions ( $\Psi$ ) which we parallelize over all DWI in our implementation.

**$\Psi$  Orthonormal.** In the case that  $\Psi$  is orthonormal, *i.e.*  $\Psi^{\top}\Psi = \Psi\Psi^{\top} = I$  then (*P1mat*) can be simplified to (5) after noticing:

$$\|\Gamma C \Psi^{\top} - S\|_F^2 = \|\Gamma C \Psi^{\top} \Psi - S\Psi\|_F^2 = \|\Gamma C - \hat{S}\|_F^2 \quad (38)$$

which can be solved using traditional sparse solvers.

**$\Psi$  Separable Tensor Product.** In the case that  $\Psi$  can be separated into a 3D tensor product  $\Psi = \Psi_x \otimes \Psi_y \otimes \Psi_z$ , the complexity of multiplication can be simplified another degree, in the same vein as the decrease in complexity we gained from using  $\Phi = \Gamma \otimes \Psi$ . Furthermore, if we consider DSI acquisition where q-space measurements are acquired in a grid over  $\mathbb{R}^3$ , and assume we can represent these measurements over a separable basis over each dimension, then we can take  $\Gamma = \Gamma_{q_x} \otimes \Gamma_{q_y} \otimes \Gamma_{q_z}$  and  $\Phi$  becomes a 6-tensor.

Algorithm	Standard	Kronecker
OMP	$k^2 + kGV + GVN_{\Gamma}N_{\Psi}$	$k^2 + kGV + GVN_{\Gamma} + VN_{\Gamma}N_{\Psi}$
OMP-PGD	–	$TG \mathcal{I}^k  \mathcal{J}^k  + TGV \mathcal{J}^k  +  \mathcal{J}^k N_{\Gamma}N_{\Psi}$
ADMM	$(GV)^2N_{\Gamma}N_{\Psi} + GV(N_{\Gamma}N_{\Psi})^2$	$3(GN_{\Gamma}N_{\Psi} + GVN_{\Psi}) + GV$
DADMM	$(GV)^2N_{\Gamma}N_{\Psi}$	$3(GN_{\Gamma}N_{\Psi} + GVN_{\Psi}) + 2GV$
FISTA	$(N_{\Gamma}N_{\Psi})^2 + GVN_{\Gamma}N_{\Psi}$	$2(GN_{\Gamma}N_{\Psi} + GVN_{\Psi})$

**Table 2:** Complexity comparison at iteration  $k$ . For Kron-OMP-PGD,  $T$  is the number of sub-iterations of PGD.

#### 4.6. Choice of Spatial-Angular Dictionaries

In this work we consider fixed spatial and angular dictionaries for HARDI data. For the choice of angular dictionary  $\Gamma$ , we consider the over-complete Spherical Ridgelet (SR) basis [40], which has been shown to sparsely model HARDI signals while ODFs are sparsely modeled with spherical wavelets (SW) counterparts (featured as the angular atom in Figure 4). With order  $L = 2$  and 4, the SR dictionary contains  $N_\Gamma = 210$  and  $N_\Gamma = 1169$  atoms, respectively. We used both amounts of atoms for the small 2D  $50 \times 50$  phantom dataset and found roughly identical results suggesting that order  $L = 2$  supply enough atoms if the number of gradients is below 210. This reduces computation significantly. In our previous work [69] we compared the performance of SR with the spherical harmonic (SH) basis which has been shown in prior work [40] to not exude sparse signals and so we do not repeat this comparison in the current work.

For the choice of spatial dictionary  $\Psi$ , the spatial wavelet transform is one popular basis for which natural images and structural MRI volumes are considered sparse. In our previous work [69] we compared the performance of Daubechies wavelets and Haar wavelets and concluded that Daubechies resulted in a boundary smoothing between isotropic and anisotropic regions which was not indicative of the more abrupt boundaries that real HARDI data exhibits. Haar outperformed Daubechies in reconstruction error due in part by HARDI’s more rigid boundaries and piece-wise consistencies, a spatial feature which have been reason for using total-variation penalties in many other reconstruction methods. For this reason, we do not consider Daubechies in this work. In addition to Haar, we consider the spatial Curvelets dictionary [63] (featured as the spatial atom in Figure 4) which, in addition to variations in position and scale, offers directional variations which may be useful for sparsely modeling the naturally directional HARDI fiber tracts regions. An important criteria for our spatial basis choices is that these spatial bases are tight frames which has important theoretical implications for CS and offers computational simplification (as discussed in Section 4.5). They additionally have fast transform implementations which also reduce computational complexity. Finally, to compare our formulation to state-of-the-art voxel-wise angular sparse coding, we can simply choose  $\Psi$  to be the identity  $I$ . To organize our dictionary choices, we use a spatial-angular  $\Psi$ - $\Gamma$  labeling Haar-SR, Curve-SR, I-SR for Haar Wavelets, Curvelets, and the Identity, respectively, combined with Spherical Ridgelets (SR).

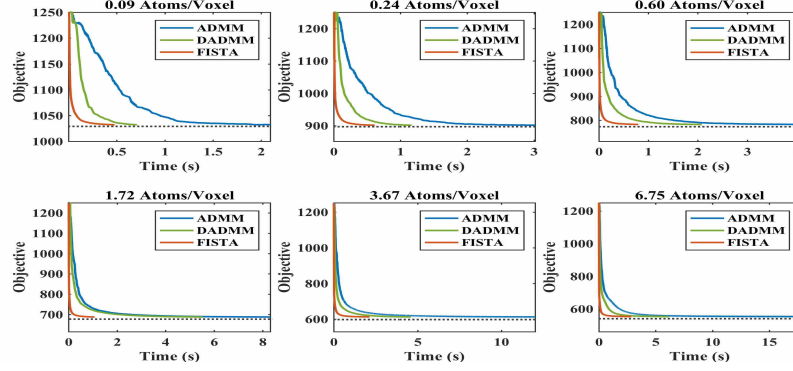
### 5. Experiments

#### 5.1. Data

In this paper the type of data we experiment on is single-shell HARDI, though as we emphasized earlier, our framework and algorithms can be applied to any type of dMRI acquisition protocol with specific choice of angular basis  $\Gamma$ . We experimented on a phantom and real HARDI brain dataset. We applied our methods on the ISBI 2013 HARDI Reconstruction Challenge Phantom dataset, a  $V = 50 \times 50 \times 50$  volume consisting of 20 phantom fibers crossing intricately within an inscribed sphere, measured with  $G = 64$  gradient directions ( $\text{SNR} = 30$ ). Our initial experiments test on a 2D  $50 \times 50$  slice of this data for simplification. The real HARDI brain dataset consists of a  $V = 112 \times 112 \times 65$  volume with  $G = 127$  gradient directions. We conducted experiments on the core white matter brain region of size  $V = 60 \times 60 \times 30$ .

#### 5.2. Kronecker Algorithm Comparison

In this section we compare the computational time performance of each of the proposed Kronecker l1 algorithms, Kron-ADMM, Kron-DADMM, and Kron-FISTA on a 2D  $50 \times 50$  slice of phantom data for various values of  $\lambda$  using Haar-SR. For our experiment, we ran Kron-FISTA until a very small error of  $10^{-8}$  was reached. The objective value obtained was then taken to be a rough ground truth minimum. We then tested each of Kron-ADMM, Kron-DADMM, and Kron-FISTA and recorded the time it took to reach a relative error of  $10^{-4}$  from the known minimum. Fig. 6 reports the objective value decent of each algorithm for various sparsity levels associated to choices of  $\lambda$ . Kron-FISTA appears to be the fastest algorithm in all cases, followed by Kron-DADMM. The superior performance of DADMM over ADMM is consistent with the findings of [70]. The average time in seconds per iteration of the three methods over the 6 sparsity levels are 0.0043, 0.0037, and 0.0058 for Kron-ADMM, Kron-DADMM, and Kron-FISTA, respectively. Even though each iteration of Kron-FISTA takes longer, the gain in speed comes from the fact that Kron-FISTA requires much fewer number of iterations. With these results, we henceforth use Kron-FISTA for all subsequent experiments.



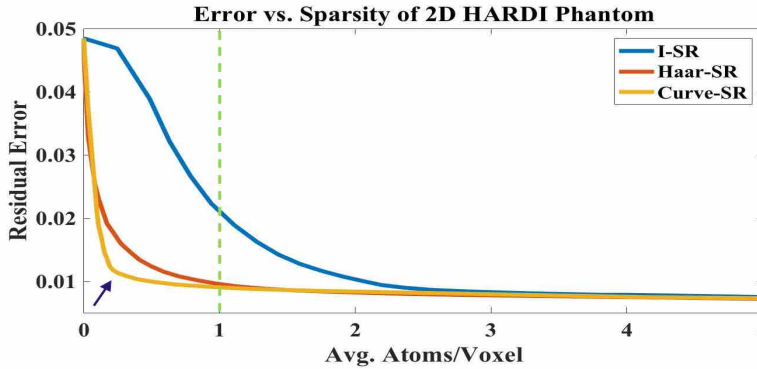
**Figure 6:** Comparison of time for completion of Kron-ADMM, Kron-DADMM, and Kron-FISTA on a 2D  $50 \times 50$  phantom HARDI data using Haar-SR for various sparsity levels. Kron-FISTA consistently reaches the known minimum objective in the least amount of time.

### 5.3. Sparsity Results

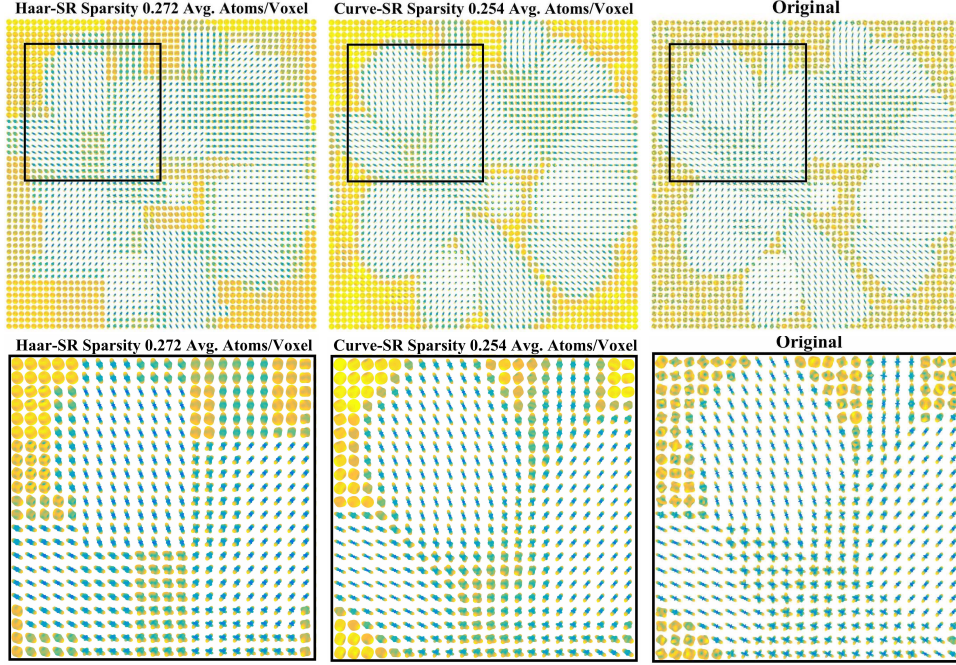
In this section we compare our spatial-angular sparse coding to state-of-the-art angular sparse coding on phantom 2D and real 3D HARDI brain data to show drastic decreases in global sparsity while maintaining accurate reconstruction.

#### 5.3.1. Phantom HARDI Data

The first experiment is tested on the  $50 \times 50$  phantom data slice. We ran Kron-FISTA for various values of  $\lambda$  for Haar-SR, Curve-SR and I-SR. In Fig. 7 we show the results of reconstruction residual error calculated as the residual  $0.5\|s^* - s\|_2/GV$  vs. global sparsity levels in terms of avg. number of atoms per voxel (K/V). The ideal reconstruction will have a very low residual error at a very high sparsity, i.e. low number of avg. atoms/voxel, occupying the lower left-hand corner of our plot. We can see that in this range, Curve-SR outperforms Haar-SR while I-SR is unable to perform at this level. Reconstruction of I-SR for various sparsity levels are visualized in Figure 2. In comparison, Fig. 8 displays the sparse reconstruction of Haar-SR and Curve-SR, each with global sparsity levels of around 0.25 avg. atoms/voxel. We can see that Curve-SR provides a smoother more accurate reconstruction than the expectedly boxy reconstruction of Haar-SR at this very high sparsity level. Our spatial-angular sparse coding can reconstruct accurate signals with much fewer atoms than the state-of-the-art.



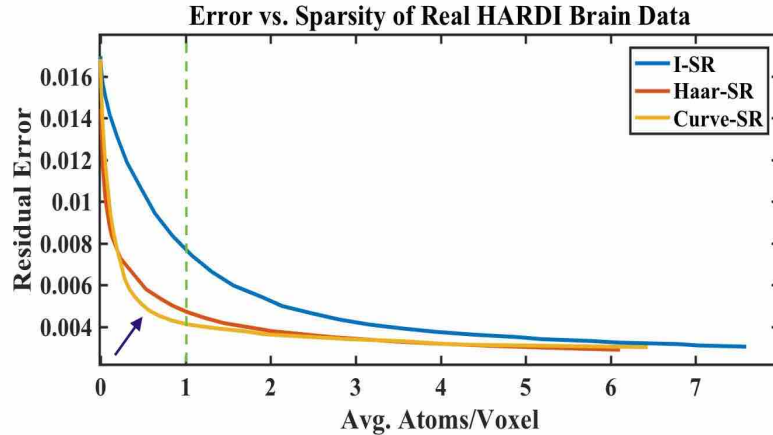
**Figure 7:** Error vs. global sparsity levels for I-SR, Haar-SR, and Curve-SR on 2D phantom HARDI data. Curve-SR outperforms Haar-SR for low sparsity levels while I-SR has very high relative reconstruction error. Our finding of I-SR requiring 6-8 atoms per voxel for accurate reconstruction is consistent with previous findings [46, 39].



**Figure 8:** Results of the proposed spatial-angular sparse coding using Kron-FISTA for Haar-SR and Curve-SR at very high sparsity level of  $\sim 0.25$  avg. atoms/voxel compared to original signal. Curve-SR outperforms Haar-SR in this high sparsity range due to its additional directionality. We can see a much better reconstruction compared to I-SR in Figure 2. This shows that we can achieve accurate reconstruction with global sparsity levels well below 1 atom/voxel.

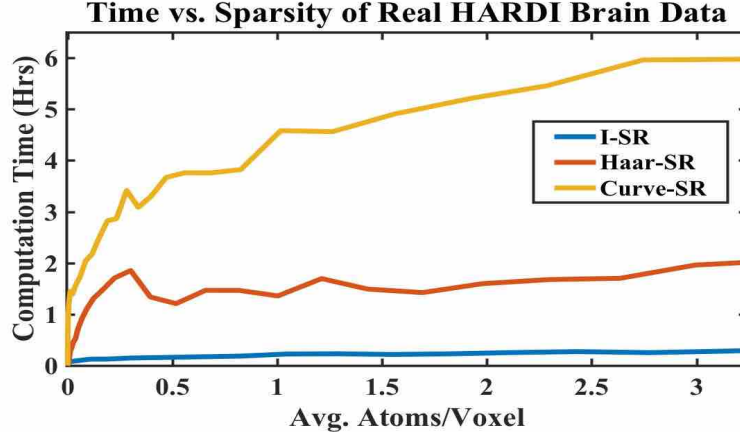
### 5.3.2. Real HARDI Brain Data

We repeated this same analysis on real 3D HARDI brain data. Fig. 9 presents the reconstruction error vs. sparsity results for I-SR, Haar-SR, and Curve-SR showing again that curvelets outperforms Haar for high sparsity levels between 0.5-3 avg. atoms/voxel. As expected and consistent with our phantom data experiment, the state-of-the-art I-SR has comparable reconstruction error in the range of 6-8 avg. atoms/voxel.



**Figure 9:** Error vs sparsity on real 3D HARDI brain data. Curve-SR provides good reconstruction error with the fewest atoms, in the range of 0.5 to 3 avg. atoms/voxel. I-SR error is much larger in this sparsity range and only comparable in the predicted range of 6-8 avg. atoms/voxel, consistent with the previously reported [46, 39].

Fig. 10 presents the computation time for each choice of dictionary for the desired sparsity range of 0.5-3 avg. atoms/voxel. We can achieve efficient completion times on this large-scale data set of size  $|S| = VG = 13,716,000$  using a MATLAB implementation. It is evident that the larger the dictionary



**Figure 10:** Computation time of Kron-FISTA for real 3D HARDI brain data for various sparsity levels.

the longer the computation time, where 3D curvelets and Haar wavelets have redundancy factors of 1.327 and 1.012, respectively, for our chosen set of parameters. In addition, the larger time discrepancy between curvelets and Haar wavelets is due to the efficient internal MATLAB implementation of wavelets compared to the external curvelets toolbox. Fig. 11 shows the quality of reconstruction of I-SR, Haar-SR, and Curve-SR compared to the original signal for the high sparsity level of  $\sim 0.5$  avg. atom/voxel. Haar-SR presents boxy regions while Curve-SR maintains a smoother reconstruction with a preservation of smaller detailed fiber tract regions. I-SR is unable to model intricate fiber regions and is forced to set most voxels to zero atoms. The proposed method can achieve much higher sparsity levels than the state-of-the-art and accurate reconstructions using less than 1 atom/voxel.

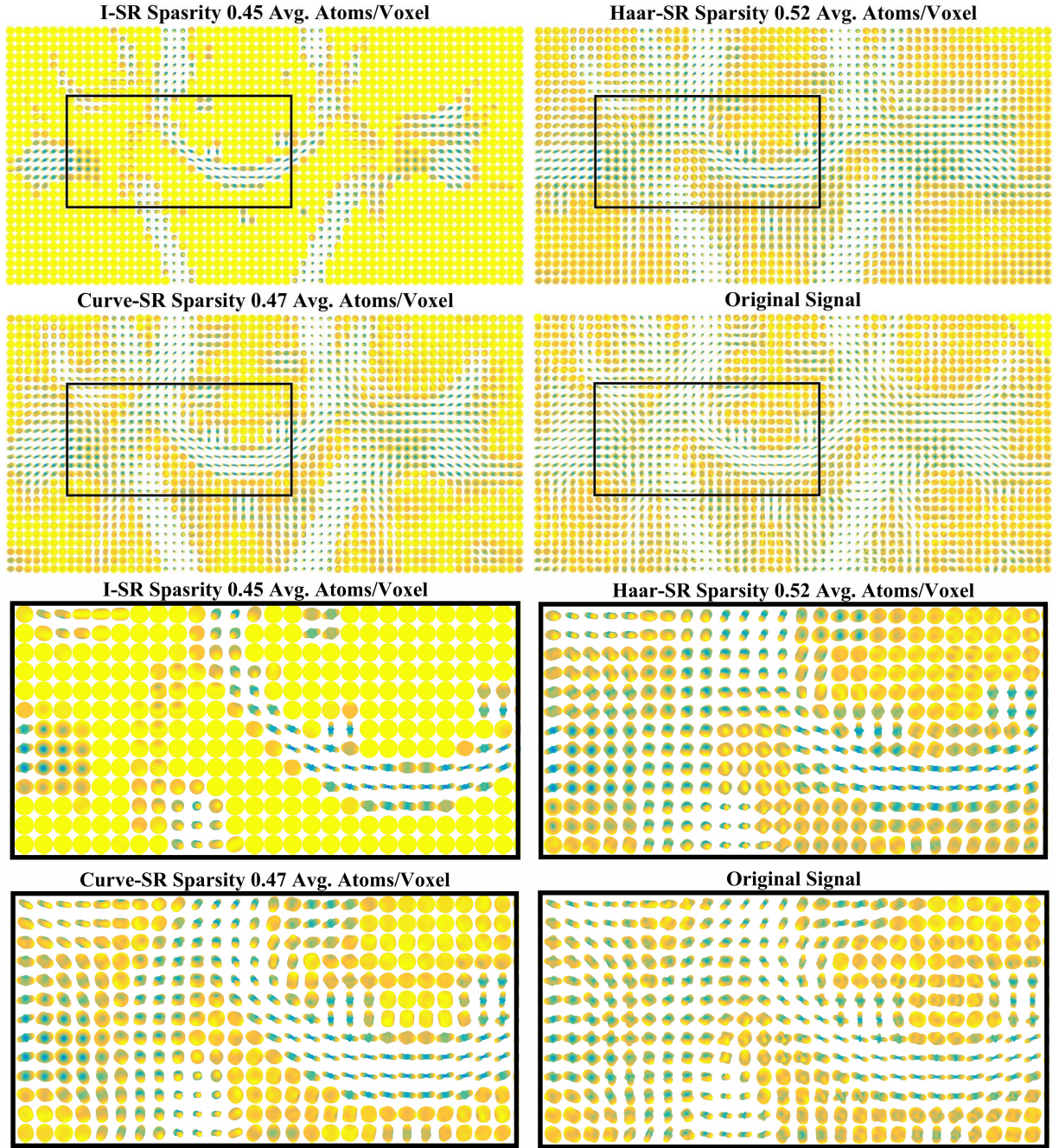
## 6. Discussion and Conclusion

In this work, we have proposed three algorithmic strategies to address sparse coding problems for the specific structure and size and dMRI data. The first, Kron-OMP-PGD, is an improvement on the existing Kron-OMP, and may be faster for applications of relatively low dimension as might be suitable for Kron-OMP, but when applied to large-scale HARDI data, it still remains largely intractable because of the high number of iterations required. We then derived a Kron-ADMM and Kron-FISTA to compare against our previously proposed Kron-DADMM and showed that Kron-FISTA was the fastest of the three and very efficient for large-scale HARDI data. In future work, we will investigate other efficient LASSO active set methods such as the recent ORacle Guided Elastic Net (ORGEN) [72].

In terms of dictionaries, it is evident from our experiments that curvelets represents the spatial texture of HARDI data more compactly than Haar wavelets due to its additional directionality and smoothness. In our future work, we will investigate other fixed spatial and angular dictionaries which may offer sparser reconstructions, such as shearlets. In addition, we aim to develop an efficient joint spatial-angular dictionary learning algorithm to learn spatial and angular sparsifying dictionaries directly from dMRI data. In general, our spatial-angular sparse coding can be applied to other dMRI data with appropriate choices of angular bases as well as general problems that exhibit separable dictionaries. Our spatial angular representation may have novel applications in other areas of dMRI processing such as denoising or fiber tract segmentation. In addition, we will investigate imposing a single global ODF non-negativity constraint following our previous work in [13] and global feature extraction following [73].

Overall, we have demonstrated that we can achieve a much sparser representation of HARDI data than the state-of-the-art by modeling HARDI with a global joint spatial-angular dictionary, allowing for the possibility of sparsity below one atom per voxel. With this higher degree of sparsity, we hope to achieve a faster acceleration of HARDI than has been previously accomplished, within a unified Kronecker (k,q)-CS framework.





**Figure 11:** Results of proposed spatial-angular sparse coding on real 3D HARDI brain data using Kron-FISTA for I-SR, Haar-SR and Curve-SR at very high sparsity level of  $\sim 0.5$  avg. atoms/voxel compared to original signal. Curve-SR provides a smoother reconstruction than the boxy Haar-SR. I-SR is unable to compete at this sparsity level.

## References

- [1] V. Wedeen, P. Hagmann, W. Tseng, T. Reese, R. Weisskoff, Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging, *Magnetic Resonance in Medicine* 54 (6) (2005) 1377–1386.
- [2] P. Basser, J. Mattiello, D. LeBihan, Estimation of the effective self-diffusion tensor from the NMR spin echo, *Journal of Magnetic Resonance* 103 (3) (1994) 247–254.
- [3] D. Tuch, Q-ball imaging, *Magnetic Resonance in Medicine* 52 (6) (2004) 1358–1372.
- [4] Y.-C. Wu, A. L. Alexander, Hybrid diffusion imaging, *NeuroImage* 36 (3) (2007) 617–629.
- [5] K. Setsompop, J. Cohen-Adad, B. Gagoski, T. Raij, A. Yendiki, B. Keil, V. J. Wedeen, L. L. Wald, Improving diffusion MRI using simultaneous multi-slice echo planar imaging, *Neuroimage* 63 (1) (2012) 569–580.
- [6] T. G. Reese, T. Benner, R. Wang, D. A. Feinberg, V. J. Wedeen, Halving imaging time of whole brain diffusion spectrum imaging and diffusion tractography using simultaneous image refocusing in EPI, *Journal of Magnetic Resonance Imaging* 29 (3) (2009) 517–522.
- [7] X. Shi, X. Ma, W. Wu, F. Huang, C. Yuan, H. Guo, Parallel imaging and compressed sensing combined framework for accelerating high-resolution diffusion tensor imaging using inter-image correlation, *Magnetic resonance in medicine* 73 (5) (2015) 1775–1785.
- [8] D. L. Donoho, M. Elad, V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. on Information Theory* 52 (1) (2006) 6–18.
- [9] M. Lustig, D. Donoho, J. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, *Magnetic Resonance in Medicine* 58 (6) (2007) 1182–1195.
- [10] L. Ning, et al., Sparse reconstruction challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use?, *Medical Image Analysis* 26 (1) (2015) 316–331.
- [11] M. Mani, M. Jacob, A. Guidon, V. Magnotta, J. Zhong, Acceleration of high angular and spatial resolution diffusion imaging using compressed sensing with multichannel spiral data, *Magnetic Resonance in Medicine* 73 (1) (2015) 126–138.
- [12] E. Schwab, B. Afsari, R. Vidal, Estimation of non-negative ODFs using eigenvalue distribution of spherical functions, in: *Medical Image Computing and Computer Assisted Intervention*, Vol. 7511, 2012, pp. 322–330.
- [13] S. Wolfers, E. Schwab, R. Vidal, Nonnegative ODF estimation via optimal constraint selection, in: *IEEE International Symposium on Biomedical Imaging*, 2014, pp. 734–737.
- [14] C. Ye, Fiber orientation estimation using nonlocal and local information, in: *Medical Image Computing and Computer Assisted Intervention*, Springer, 2016, pp. 97–105.
- [15] A. Goh, C. Lenglet, P. Thompson, R. Vidal, Estimating orientation distribution functions with probability density constraints and spatial regularity, in: *Medical Image Computing and Computer Assisted Intervention*, Vol. 5761, 2009, pp. 877–885.
- [16] B. Yoldemir, M. Bajammal, R. Abugharbieh, Dictionary Based Super-Resolution for Diffusion MRI, in: *MICCAI Workshop on Computational Diffusion MRI*, Springer, 2014, pp. 203–213.
- [17] Y. Ouyang, Y. Chen, Y. Wu, H. M. Zhou, Total variation and wavelet regularization of orientation distribution functions in diffusion MRI, *Inverse Problems and Imaging* 7 (2) (2013) 565–583.
- [18] J. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Transactions on Information Theory* 50 (10) (2004) 2231–2242.
- [19] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends in Machine Learning* 3 (1) (2010) 1–122.
- [20] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences* 2 (1) (2009) 183–202.
- [21] S. Merlet, J. Cheng, A. Ghosh, R. Deriche, Spherical polar Fourier EAP and ODF reconstruction via compressed sensing in diffusion MRI, in: *IEEE International Symposium on Biomedical Imaging*, IEEE, 2011, pp. 365–371.
- [22] R. Aranda, A. Ramirez-Manzanares, M. Rivera, Sparse and Adaptive Diffusion Dictionary (SADD) for recovering intra-voxel white matter structure, *Medical Image Analysis* 26 (1) (2015) 243–255.
- [23] B. Bilgic, K. Setsompop, J. Cohen-Adad, A. Yendiki, L. L. Wald, E. Adalsteinsson, Accelerated diffusion spectrum imaging with compressed sensing using adaptive dictionaries, *Magnetic Resonance in Medicine* 68 (6) (2012) 1747–1754.
- [24] S. Merlet, E. Caruyer, R. Deriche, Parametric dictionary learning for modeling EAP and ODF in diffusion MRI, in: *Medical Image Computing and Computer Assisted Intervention*, Springer, 2012, pp. 10–17.

- [25] S. Merlet, E. Caruyer, A. Ghosh, R. Deriche, A computational diffusion MRI and parametric dictionary learning framework for modeling the diffusion signal and its features, *Medical Image Analysis* 17 (7) (2013) 830–843.
- [26] J. Sun, Y. Xie, W. Ye, J. Ho, A. Entezari, S. J. Blackband, B. C. Vemuri, Dictionary learning on the manifold of square root densities and application to reconstruction of diffusion propagator fields, in: *International Conference on Information Processing in Medical Imaging*, Springer, 2013, pp. 619–631.
- [27] A. Gramfort, C. Poupon, M. Descoteaux, Denoising and fast diffusion imaging with physically constrained sparse dictionary learning, *Medical Image Analysis* 18 (1) (2014) 36–49.
- [28] J. Cheng, T. Jiang, R. Deriche, D. Shen, P.-T. Yap, Regularized spherical polar fourier diffusion MRI with optimal dictionary learning, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2013, pp. 639–646.
- [29] J. Cheng, D. Shen, P.-T. Yap, P. J. Basser, Tensorial spherical polar fourier diffusion mri with optimal dictionary learning, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 174–182.
- [30] S. Merlet, R. Deriche, Compressed sensing for accelerated EAP recovery in diffusion MRI, in: *MICCAI*, 2010, pp. Page–14.
- [31] M. Menzel, E. Tan, K. Khare, J. Sperl, K. King, X. Tao, C. Hardy, L. Marinelli, Accelerated diffusion spectrum imaging in the human brain using compressed sensing, *Magnetic Resonance in Medicine* 66 (5) (2011) 1226–1233.
- [32] S. Merlet, R. Deriche, Continuous diffusion signal, EAP and ODF estimation via compressive sensing in diffusion MRI, *Medical Image Analysis* 17 (5) (2013) 556–572.
- [33] M. Paquette, S. Merlet, G. Gilbert, R. Deriche, M. Descoteaux, Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging, *Magnetic Resonance in Medicine* 73 (1) (2015) 401–416.
- [34] J. Cheng, S. Merlet, E. Caruyer, A. Ghosh, T. Jiang, R. Deriche, Compressive sensing ensemble average propagator estimation via l1 spherical polar fourier imaging, in: *MICCAI Workshop on Computational Diffusion MRI*, 2011.
- [35] Y. Rathi, O. Michailovich, K. Setsompop, S. Bouix, M. Shenton, C.-F. Westin, Sparse multi-shell diffusion imaging, in: *Medical Image Computing and Computer Assisted Intervention*, Springer, 2011, pp. 58–65.
- [36] J. M. Duarte-Carvajalino, C. Lenglet, J. Xu, E. Yacoub, K. Ugurbil, S. Moeller, L. Carin, G. Sapiro, Estimation of the CSA-ODF using Bayesian compressed sensing of multi-shell HARDI, *Magnetic Resonance in Medicine* 72 (5) (2014) 1471–1485.
- [37] A. Daducci, E. J. Canales-Rodríguez, H. Zhang, T. B. Dyrby, D. C. Alexander, J.-P. Thiran, Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data, *NeuroImage* 105 (2015) 32–44.
- [38] O. Michailovich, Y. Rathi, Fast and accurate reconstruction of HARDI data using compressed sensing, in: *Medical Image Computing and Computer Assisted Intervention*, Springer, 2010, pp. 607–614.
- [39] O. Michailovich, Y. Rathi, On approximation of orientation distributions by means of spherical ridgelets, *IEEE Transactions on Image Processing* 19 (2) (2010) 461–477.
- [40] A. Tristán-Vega, C.-F. Westin, Probabilistic ODF estimation from reduced HARDI data with sparse regularization, in: *Medical Image Computing and Computer Assisted Intervention*, Springer, 2011, pp. 182–190.
- [41] J. M. Duarte-Carvajalino, C. Lenglet, K. Ugurbil, S. Moeller, L. Carin, G. Sapiro, A framework for multi-task Bayesian compressive sensing of DW-MRI, in: *MICCAI Workshop on Computational Diffusion MRI*, 2012, pp. 1–13.
- [42] I. B. Alaya, M. Jribi, F. Ghorbel, T. Kraiem, A Novel Geometrical Approach for a Rapid Estimation of the HARDI Signal in Diffusion MRI, in: *International Conference on Image and Signal Processing*, Springer, 2016, pp. 253–261.
- [43] B. A. Landman, J. A. Bogovic, H. Wan, F. E. Z. ElShahaby, P.-L. Bazin, J. L. Prince, Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI, *NeuroImage* 59 (3) (2012) 2175–2186.
- [44] D. Kuhnt, M. H. Bauer, J. Egger, M. Richter, T. Kapur, J. Sommer, D. Merhof, C. Nimsy, Fiber tractography based on diffusion tensor imaging compared with high-angular-resolution diffusion imaging with compressed sensing: initial experience, *Neurosurgery* 72 (0 1) (2013) 165.
- [45] D. Kuhnt, M. H. A. Bauer, J. Sommer, D. Merhof, C. Nimsy, Optic radiation fiber tractography in glioma patients based on high angular resolution diffusion imaging with compressed sensing compared with diffusion tensor imaging - initial experience, *PLoS ONE* 8 (7) (2013) 1–7.
- [46] O. Michailovich, Y. Rathi, On approximation of orientation distributions by means of spherical ridgelets, in: *IEEE International Symposium on Biomedical Imaging*, IEEE, 2008, pp. 939–942.
- [47] Y. Ouyang, Y. Chen, Y. Wu, Vectorial total variation regularisation of orientation distribution functions in diffusion weighted MRI, *International Journal of Bioinformatics Research and Applications* 10 (1) (2014) 110–127.

- [48] C. Ye, J. L. Prince, Probabilistic tractography using Lasso bootstrap, *Medical Image Analysis* 35 (2017) 544–553.
- [49] O. Michailovich, Y. Rath, S. Dolui, Spatially regularized compressed sensing for high angular resolution diffusion imaging, *IEEE Transactions on Medical Imaging* 30 (5) (2011) 1100–1115.
- [50] A. Gramfort, C. Poupon, M. Descoteaux, Sparse DSI: Learning DSI structure for denoising and fast imaging, in: *Medical Image Computing and Computer Assisted Intervention*, Springer, 2012, pp. 288–296.
- [51] W. Ye, B. C. Vemuri, A. Entezari, An over-complete dictionary based regularized reconstruction of a field of ensemble average propagators, in: *IEEE International Symposium on Biomedical Imaging*, Springer, 2012, pp. 940–943.
- [52] Y. Rath, O. Michailovich, F. Laun, K. Setsompop, P. E. Grant, C.-F. Westin, Multi-shell diffusion signal recovery from sparse measurements, *Medical Image Analysis* 18 (7) (2014) 1143–1156.
- [53] A. Auría, A. Daducci, J.-P. Thiran, Y. Wiaux, Structured sparsity for spatially coherent fibre orientation estimation in diffusion MRI, *NeuroImage* 115 (2015) 245–255.
- [54] S. P. Awate, E. V. DiBella, Compressed sensing HARDI via rotation-invariant concise dictionaries, flexible K-space undersampling, and multiscale spatial regularity, in: *IEEE International Symposium on Biomedical Imaging*, IEEE, 2013, pp. 9–12.
- [55] J. Cheng, D. Shen, P. J. Basser, P. T. Yap, Joint 6D kq space compressed sensing for accelerated high angular resolution diffusion MRI, in: *Information Processing in Medical Imaging*, Springer, 2015, pp. 782–793.
- [56] J. Sun, E. Sakhanev, A. Entezari, B. C. Vemuri, Leveraging eap-sparsity for compressed sensing of ms-hardi in (k,q)-space, in: *Information Processing in Medical Imaging*, Springer, 2015, pp. 375–386.
- [57] D. McClymont, I. Teh, H. J. Whittington, V. Grau, J. E. Schneider, Prospective acceleration of diffusion tensor imaging with compressed sensing using adaptive dictionaries, *Magnetic Resonance in Medicine*.
- [58] F. Knoll, J. G. Raya, R. O. Halloran, S. Baete, E. Sigmund, R. Bammer, T. Block, R. Otazo, D. K. Sodickson, A model-based reconstruction for undersampled radial spin-echo DTI with variational penalties on the diffusion tensor, *NMR in Biomedicine* 28 (3) (2015) 353–366.
- [59] L. Ning, K. Setsompop, O. V. Michailovich, N. Makris, M. E. Shenton, C.-F. Westin, Y. Rath, A joint compressed-sensing and super-resolution approach for very high-resolution diffusion imaging, *NeuroImage* 125 (2016) 386–400.
- [60] T.-C. Chao, J.-y. G. Chiou, S. E. Maier, B. Madore, Fast diffusion imaging with high angular resolution, *Magnetic Resonance in Medicine* (2016) n/a–n/doi:10.1002/mrm.26163.  
URL <http://dx.doi.org/10.1002/mrm.26163>
- [61] S. Yin, X. You, W. Xue, B. Li, Y. Zhao, X.-Y. Jing, P. S. Wang, Y. Tang, A Unified Approach for Spatial and Angular Super-Resolution of Diffusion Tensor MRI, in: *Chinese Conference on Pattern Recognition*, Springer, 2016, pp. 312–324.
- [62] A. C. Yang, M. Kretzler, S. Sudarski, V. Gulani, N. Seiberlich, Sparse reconstruction techniques in magnetic resonance imaging: Methods, applications, and challenges to clinical adoption, *Investigative radiology* 51 (6) (2016) 349–364.
- [63] E. Candes, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, *Multiscale Modeling & Simulation* 5 (3) (2006) 861–899.
- [64] C. F. Caiafa, A. Cichocki, Computing sparse representations of multidimensional signals using kronecker bases, *Neural Computation* 25 (1) (2013) 186–220.
- [65] S. Hawe, M. Seibert, M. Kleinsteuber, Separable dictionary learning, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 438–445.
- [66] M. F. Duarte, R. G. Baraniuk, Kronecker compressive sensing, *IEEE Transactions on Image Processing* 21 (2) (2012) 494–504.
- [67] C. F. Caiafa, F. Pestilli, Sparse multiway decomposition for analysis and modeling of diffusion imaging and tractography, *arXiv preprint arXiv:1505.07170*.
- [68] Y. Rivenson, A. S., Compressed imaging with a separable sensing operator, *IEEE Signal Processing Letters* 16 (6) (2009) 449–452.
- [69] E. Schwab, R. Vidal, N. Charon, Spatial-Angular Sparse Coding for HARDI, in: *Medical Image Computing and Computer Assisted Intervention*, Springer, 2016, pp. 475–483.
- [70] H. R. Goncalves, Accelerated sparse coding with overcomplete dictionaries for image processing applications, Ph.D. thesis (2015).
- [71] N. Qi, Y. Shi, X. Sun, B. Yin, TenSR: Multi-dimensional tensor sparse representation, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5916–5925.

- [72] C. You, C.-G. Li, D. Robinson, R. Vidal, Oracle based active set algorithm for scalable elastic net subspace clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [73] E. Schwab, H. E. Cetingül, B. Afsari, M. A. Yassa, R. Vidal, Rotation invariant features for HARDI, in: Information Processing in Medical Imaging, 2013, pp. 322–330.